

001043

UNIVERSIDAD DEL VALLE DE GUATEMALA
FACULTAD DE CIENCIAS Y HUMANIDADES
MAESTRIA EN TECNOLOGIA Y CIENCIAS DE LA
COMPUTACION

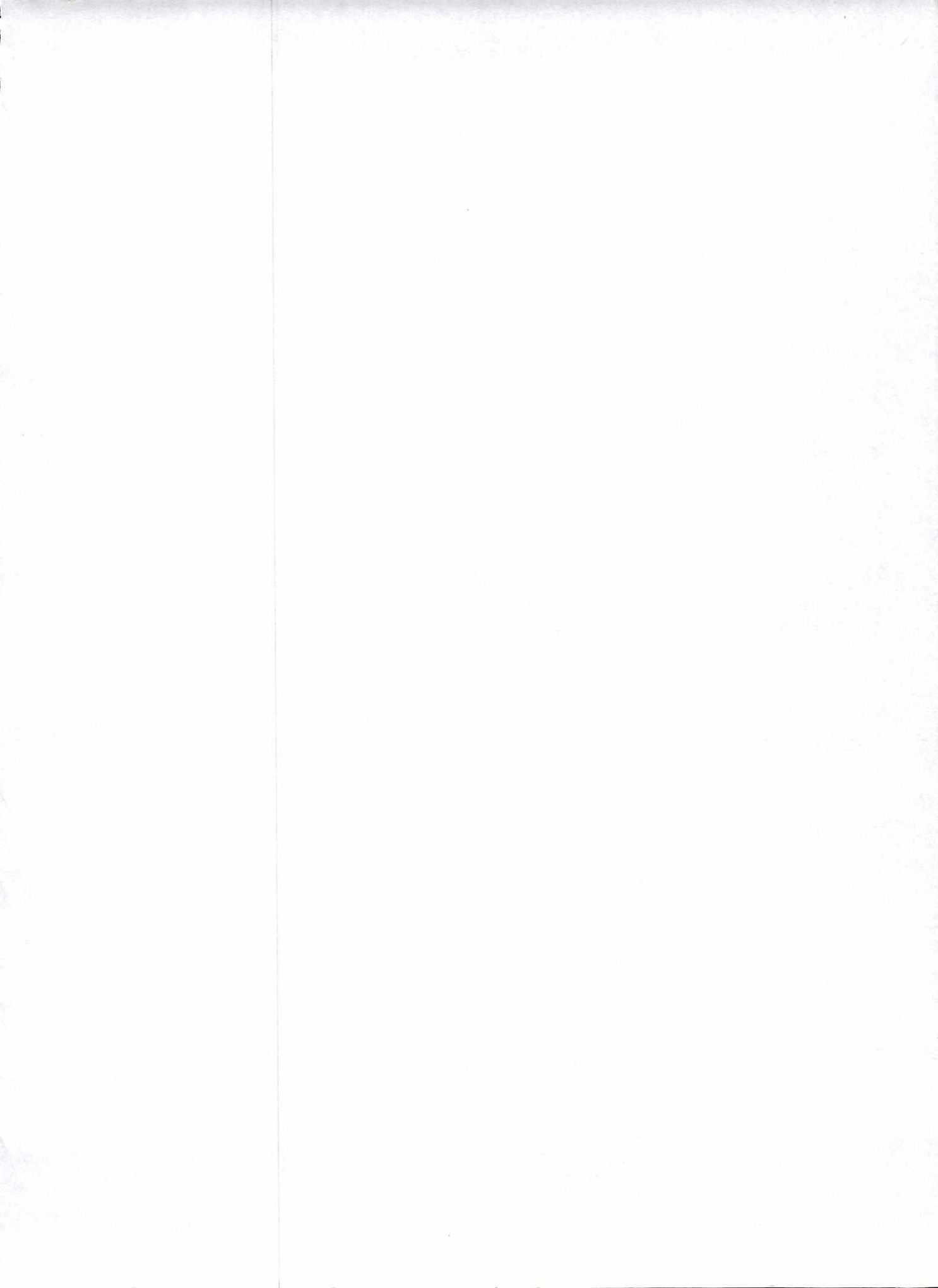


MODELO COMPUTARIZADO DE ANALISIS Y EVALUACION
DE DATOS, CON RESULTADOS ESTADISTICOS

VICTOR HUGO MORALES MOLINA

Guatemala

2001



MODELO COMPUTARIZADO DE ANALISIS Y EVALUACION
DE DATOS CON RESULTADOS ESTADISTICOS



UNIVERSIDAD DEL VALLE DE GUATEMALA
FACULTAD DE CIENCIAS Y HUMANIDADES
MAESTRIA EN TECNOLOGIA Y CIENCIAS DE LA
COMPUTACION

MODELO COMPUTARIZADO DE ANALISIS Y EVALUACION
DE DATOS, CON RESULTADOS ESTADISTICOS

VICTOR HUGO MORALES MOLINA

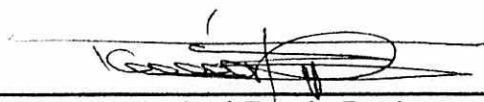
Trabajo de Graduación para optar al grado académico de
Maestría en Tecnología y Ciencias de la Computación

Guatemala

2001

Vo. Bo.

(f)



Licenciado José Tomás Barrientos
Asesor

Tribunal:

(f)



Licenciado José Tomás Barrientos

(f)



Ingeniero Arturo Rivera

(f)



Ingeniera Patricia Castillo

Fecha de aprobación: 30 de julio de 2001.

DEDICATORIA

A Dios por tenerme aún con vida.

A mi asesor el Dr. Tomás Barrientos, por creer que sí se puede y por haberme tolerado en sus clases, de donde surgió la idea para el presente trabajo profesional.

A la Licda. Liseth Flores Roa, M.A, por enseñarme a escalar las cumbres de la metodología e investigación, con un espíritu altruista e infatigable.

TABLA DE CONTENIDO

| | Páginas |
|---------------------------------------------------------------------------------|---------|
| I. INTRODUCCIÓN..... | 1 |
| II. JUSTIFICACION..... | 3 |
| III. MARCO REFERENCIAL SOBRE ANALISIS Y EVALUACION | 7 |
| A. Principios de Análisis y Evaluación de Datos..... | 7 |
| 1. Análisis de Datos..... | 7 |
| 2. Definición General de Datos..... | 8 |
| 3. Evaluación de Datos..... | 8 |
| B. Conceptos Generales del Aspecto Estadístico..... | 9 |
| 1. Orígenes de la Estadística..... | 9 |
| 2. Síntesis Histórica de las estadísticas..... | 12 |
| 3. Algunas Definiciones de Estadística y Métodos Estadísticos..... | 12 |
| 4. Métodos estadísticos..... | 14 |
| 5. Tipos de Estadísticas..... | 14 |
| 6. Conceptos de Procesos Utilizados en el análisis estadístico de este modelo.. | 15 |
| 6.1 Distribución de Frecuencias..... | 15 |
| 6.2 Correlación..... | 16 |
| 6.3 Análisis de Varianza..... | 16 |
| 6.4 Variable..... | 16 |
| 7. Campos de la aplicación de la Estadística..... | 16 |
| 7.1 Estadística en el Campo Social..... | 17 |
| 7.2 Sistemas Computarizados y Estadística..... | 18 |
| C. Conceptos del Sistema de Cómputo..... | 20 |
| 1. El Sistema Operativo..... | 20 |
| 2. Anatomía Básica de un Sistema Operativo..... | 21 |
| 3. Red de Computadoras..... | 22 |
| a. Red de Area Local..... | 22 |
| b. Red de Tipo Anillo..... | 23 |
| c. Red de Tipo Bus..... | 24 |
| d. Red de Tipo Estrella..... | 25 |
| e. Red de Area Remota..... | 26 |
| f. Plataforma de Operaciones de una Red..... | 27 |
| 4. Generalidades de Olap..... | 29 |
| a. Reportes Estadísticos de Cubo..... | 34 |
| b. Data Mining..... | 36 |
| 5. Compatibilidad de Herramientas de Software..... | 41 |
| a. Generalidades de Excel 97..... | 41 |
| b. Generalidades de Access 97..... | 42 |
| IV. ANALISIS METODOLOGICO DEL MODELO PROPUESTO..... | 45 |
| A. Investigación de Requerimientos Básicos..... | 45 |
| 1. Análisis de Datos Basados en Hechos Reales..... | 45 |
| 2. Identificación de Requerimientos Esenciales..... | 46 |
| 3. Selección de Estrategias para Satisfacer los requerimientos..... | 46 |
| B. Técnica Administrativa de Análisis..... | 46 |

| | | |
|-------|-----------------------------------------------------------------|----|
| C. | Estudio de Campo en San José Poaquil..... | 48 |
| | 1. Empresa a la cual se analizó..... | 49 |
| | 2. Datos Relevantes de las Primeras Entrevistas..... | 49 |
| D. | Descripción de la Forma como Surge el Problema..... | 56 |
| | 1. Técnicas y Procedimientos de la Selección de la Muestra..... | 57 |
| | 2. Momento en que Surge el Problema..... | 59 |
| E. | Descripción de hardware utilizado..... | 61 |
| | 1. Computadoras..... | 61 |
| | 2. Impresoras..... | 61 |
| F. | Software Utilizado por la Empresa de Investigación..... | 62 |
| | 1. Sistema Operativo..... | 62 |
| | 2. Programas de apoyo al usuario..... | 62 |
| V. | ESTRUCTURA DEL MODELO PROPUESTO..... | 64 |
| A. | Programas necesarios en el Modelo..... | 65 |
| | 1. Sistema Operativo y Enlace de Red..... | 65 |
| | 2. Programas de Aplicación..... | 67 |
| | 3. Programas de Soporte..... | 71 |
| | 4. Generalidades de SPSS 10.0..... | 72 |
| B. | Equipo Utilizado en el Modelo..... | 76 |
| | 1. Oficina de Dirección y Análisis Estadístico..... | 76 |
| | 2. Oficina de Dirección de Control de Procesos..... | 77 |
| | 3. Oficina de Recepción y Entrada de Datos..... | 79 |
| C. | Equipo para Enlace de Red..... | 80 |
| VI. | OBSERVACIONES FINALES..... | 83 |
| VII. | BIBLIOGRAFIA..... | 85 |
| VIII. | GLOSARIO..... | 87 |
| IX. | ANEXO..... | 97 |

I. INTRODUCCION

El uso de representaciones visuales para comunicar información cuantitativa ha sido una faceta de la estadística que se ha descuidado. Afortunadamente, la amplia disponibilidad de computadores personales y software gráfico están ayudando a corregir esta situación.

La información estadística puede comunicarse más fácilmente si se organiza en tablas y se muestra en gráficas. Los procedimientos para organizar, abreviar y simplificar datos son fundamentales para los métodos estadísticos. El uso creciente de la cuantificación y el análisis estadístico en cualquiera de los campos, ha marcado dramáticamente la necesidad de organizar datos para comunicar un significado o un resultado que conlleve a la toma correcta de decisiones.

Las últimas generaciones de investigadores, en el área de desarrollo de sistemas de información, se han concentrado bastante para encontrar técnicas que permitan incrementar la productividad y el control de calidad en los procesos de elaboración de software y hoy en día la tecnología reemplaza al papel y al lápiz del técnico en la materia, por un computador, herramientas de software y metodología de software estructurado.

Estas herramientas están destinadas a ser utilizadas por todo profesional dedicado al análisis y diseño de sistemas de información, líderes de proyectos, analistas y programadores, que desean contar con los instrumentos esenciales aplicables en las labores cotidianas del área de trabajo en que se desenvuelven.

Este Modelo de Análisis y Evaluación de Datos con Resultados Estadísticos, es una respuesta al proveer al usuario final de una herramienta apropiada para estar al día en el avance tecnológico de nuestra sociedad actual. Lamentablemente muchas organizaciones no gubernamentales que apoyan a gran diversidad de comunidades de nuestro país, no cuentan con las metodologías adecuadas, equipo, programas de computación, ni el personal especializado para la formulación y evaluación de proyectos con un análisis estadístico.

Usted encontrará en este Modelo Computarizado de Análisis y Evaluación de Datos, cinco capítulos, glosario, bibliografía y anexos. El primero de ellos plantea la naturaleza y la justificación del modelo.

El segundo, describe los principios fundamentales del análisis y evaluación de datos, conceptos generales del aspecto estadístico, definiciones de los procesos estadísticos utilizados en el análisis del modelo y del sistema de cómputo.

En el capítulo tercero se hace el análisis metodológico que incluye la investigación exhaustiva de los requerimientos básicos, técnicas de análisis, investigación de campo así como el análisis de resultados, alcances y limitaciones del trabajo de campo, el cuarto describe y define la estructura del modelo propuesto.

En la parte final se incluyen algunas observaciones, las cuales pueden servir de guía a otras empresas, tomando en cuenta sus requerimientos y que éstos sean capaces de trabajar en forma "transparente" con información dispersa en varias bases de datos diferentes, ejecutadas en computadores distintos, apoyadas por diversos sistemas operativos y conectadas entre sí mediante redes de comunicación distintas.

El glosario, bibliografía y los anexos pueden servir de referencia valiosa al lector, ya que están relacionados con la evaluación de proyectos, definiciones estadísticas, conceptos del sistema de cómputo y programas sociales.

II. JUSTIFICACION

“Los Antropólogos han argumentado sobre si los humanos llegaron a ser inteligentes porque caminaron en una posición erecta, porque tuvieron que matar otros animales sin tener mandíbulas y garras poderosas necesarias para el trabajo o porque ellos aprendieron de cómo comunicarse utilizando formas complejas de articulación y vocalización. Sin embargo volviendo atrás y revisando nuestras definiciones referentes a los seres humanos como utilizadores de herramientas y lenguajes, está el concepto anterior de que éstos son utilizadores de información. La información y su uso nos permite construir y utilizar las herramientas, pero también cómo expresarnos por medio de un lenguaje”. :Kamran Parsaye:V:

Se ha dicho por varios medios, que estamos viviendo en la era de la información, pero regularmente somos golpeados por el hecho de que mucha tecnología ya existente, es: subutilizada y tiende a ser tratada en ambientes aislados o separados.

En los últimos años se han logrado muchos avances en diferentes campos que tienen importancia crítica en la tecnología de la información. Estos avances han sido significativos pero han estado fragmentados en sus diferentes áreas. Ahora bien, juntar todos esos fragmentos de información, en un marco más unificado, beneficiará a las personas que se desarrollan en diferentes disciplinas, contribuyendo también a facilitar las actividades entre seres humanos.

Uno de los problemas que resulta de la subutilización de herramientas, por falta de información, es la predicción del tamaño de un determinado proyecto; que puede ser un programa de computación, un edificio, una consultoría, capacidades de hardware para determinada compra, inclusive hasta la estimación de cuánto papel membretado y sobres hay que imprimir cuando se abre un nuevo negocio, etc. Asimismo es oportuno mencionar que en muchas ocasiones, no sólo el miedo a lo desconocido, sino también el factor económico juega un papel muy importante para definir los mínimos de operación al iniciar un evento, sin tomar en cuenta el margen de crecimiento o proyección organizacional.

Este trabajo profesional: **“MODELO COMPUTARIZADO DE ANALISIS Y EVALUACION DE DATOS CON RESULTADOS ESTADISTICOS”** nace precisamente de una situación

donde las estimaciones y predicciones sufrieron el choque de la subutilización de herramientas y el uso aislado de las mismas. Especialmente el último choque lleva a lo que en términos de procesamiento de datos se le conoce como "doble manejo de información", que da resultados con costos altos en dinero y tiempo.

Después de buscar varias empresas que permitieran implementar el presente modelo de análisis de datos, se encontró a empresa "*Soluciones de Investigación Estadística*", quienes accedieron a que se practicaran las tareas experimentales en su localidad; además de permitir utilizar sus instalaciones, también dieron el equipó de cómputo, información completa de algunos proyectos ejecutados en el municipio de San José Poaquil, del departamento de Chimaltenango, donde se les presentó una situación adversa en la cual ellos querían se practicara la mayor parte de tareas de investigación experimental del referido modelo. También dieron muchas horas para el análisis, información sobre las aplicaciones estadísticas y participación en el diseño final. Más adelante se hizo una descripción completa de su organización y de sus actividades. Para efectos de referencia, cada vez que se hable de "*Soluciones de Investigación Estadística*", se le llamará: La empresa de investigación.

En esta empresa de investigación, como se mencionó anteriormente, se les presentó una situación adversa, la cual dio origen a creer que lo que originalmente se pensó como una consultoría sencilla de análisis y resultados estadísticos y de dimensiones no muy grandes, resultó ser un servicio mucho más extenso, tanto en la cantidad de usuarios, como en los datos por analizar. Las consultorías, resultados y propuestas al usuario han funcionado adecuadamente, sin embargo se han encontrado ante la situación, de que para producir algunos resultados o reportes estadísticos, ha sido necesario dividir muchas bases de datos y sus respectivas tablas, a fin de poder aplicar diversos procedimientos estadísticos, por ejemplo: separar ciertas variables, luego volver a unir las mismas, tabularlas, analizarlas, etc., hasta llegar a un resultado final satisfactorio. En resumen, los resultados fueron los que el usuario necesitaba inicialmente, sin embargo se tuvo que recurrir al doble manejo de información y tener que depender de un especialista en sistemas de computación, además del tiempo adicional que no estaba contemplado cuando se hizo el presupuesto; todo esto redujo el margen de ganancia y aumentó las horas hombre, a fin de cumplir en el tiempo programado. No está demás mencionar que se corría mucho riesgo de sesgar la

información, al manipular datos de una tabla a otra o al tener que hacer filtros y clasificaciones de los mismos datos, si no se tenía el cuidado necesario.

Por otra parte, esas mismas bases de datos debían ser utilizadas en una investigación donde había que evaluar el impacto que han tenido los proyectos, especialmente en las comunidades en donde se llevaron a cabo. Una vez más la empresa de investigación, empezó a pensar en la situación que se había dado con anterioridad y era necesario tomar las medidas para que no sucediera igual.

Para la mayor parte de consultorías con muestras de dimensiones no tan grandes y con muchas variables, la empresa de investigación ha utilizado dos herramientas en cómputo, como son los programas Excel y Access, sin embargo se necesita algo más que una hoja electrónica y un manejador de datos con tablas relacionales, para atender adecuadamente estos casos.

Se necesita un sistema que no sólo pueda capturar la información y tabularla, sino que más inteligentemente y de manera simultánea, utilice los diferentes procedimientos estadísticos con toda la gama de parámetros que usualmente sirven para evaluar la muestra original y producir los resultados en tabulares y gráficos más funcionales; evitando así el doble manejo de información o la manipulación de ésta, que en algún momento pudiera, como se dijo anteriormente, causar sesgos en los resultados.

Esta herramienta de recursos combinados, que aplicará directamente la mayor parte de procedimientos estadísticos, debe estar instalada sobre un sistema operativo, de fácil uso por parte de los investigadores y no necesariamente obligue a recurrir a un Ingeniero en Sistemas. Por otra parte, debe proveer diferentes niveles de seguridad, ya que los datos en la mayor parte de casos son confidenciales.

En lo que se refiere al equipo a utilizar y su distribución, la empresa de investigación manifestó el deseo de tener un computador central, con la información específicamente para análisis, que contenga las bases de datos, tablas y el o los programas integrales de apoyo, a fin de compartirlos con dos estaciones locales. También debe recibir transmisiones de información, de una o más computadoras remotas, las cuales pueden ser portátiles; éstas deben transmitir los datos de las muestras recogidas en diferentes regiones y combinarlas en el servidor central, para su análisis. Aparte de ser sencillo de

operar y administrar, debe estar dentro de un presupuesto ya determinado, no muy elevado.

III. MARCO REFERENCIAL SOBRE ANALISIS Y EVALUACION

A. Principios de Análisis y Evaluación de Datos:

Normalmente cuando un investigador obtiene información, separa ésta en las partes que conforman los estudios para responder a las preguntas investigadas, así como para evaluar las hipótesis. Sin embargo, el análisis de los datos investigados, frecuentemente no proporcionan por sí mismos las respuestas a dichas preguntas que interesan más en el proyecto. Para ello se requiere del análisis y evaluación de los datos, así como su interpretación y cuando se habla de interpretación, se hace referencia a explicar y encontrar el significado de la información obtenida. Se puede decir que resulta difícil o imposible explicar los datos crudos, así que debe procederse primero al analizar los datos y luego ya se pueden interpretar y evaluar los resultados estadísticos.

1. Análisis de datos

Significa categorizar, ordenar, manipular y resumir los datos de una investigación, para dar respuesta a las preguntas planteadas. Su propósito está enfocado a reducir la información a una forma entendible e interpretable, de tal manera que las relaciones de los problemas de la investigación puedan estudiarse y evaluarse. Es pertinente mencionar que uno de los propósitos centrales de la estadística es manipular y resumir los datos numéricos y comparar los resultados obtenidos con las expectativas previstas. También al análisis de datos se le conoce como *"el proceso de extraer de las mediciones dadas, una información relevante con la cual se puede formular una descripción numérica, sumaria y comprensible"*. La interpretación hace uso de los resultados obtenidos a través del análisis y hace inferencias pertinentes a las relaciones de investigación para así dar conclusiones previas a la toma de decisiones. Esto quiere decir que el análisis y la interpretación están estrechamente ligados ya, que cuando alguien interpreta, automáticamente analiza.

2. Definición general de datos

Es la información que se da o el antecedente necesario para llegar al conocimiento de una cosa; de la cual se derivan inferencias; éstas suelen ser expresados como variables (números) o atributos (cualidades), los punteos en las pruebas y estadígrafos, como las medias y otros parámetros absolutos y relativos, porcentajes, coeficientes de correlación, etc. La palabra datos, algunas veces, designa los resultados de análisis estadístico o matemático.

3. Evaluación de datos

En la vida cotidiana, cuando las personas detectan un problema, recogen información para tomar decisiones que permitan enfrentarlo de la mejor manera posible. Del mismo modo, realizadas las acciones que se consideran oportunas en su momento, suelen reflexionar sobre el acierto o el error en que incurrieron al actuar como lo hicieron. De esta forma es como se evalúan acciones y se obtiene información para ajustar su comportamiento futuro. Algunas veces son evaluaciones teñidas de subjetividad y que se realizan, en la mayoría de los casos, con información insuficiente y mediante una reflexión asistemática.

Lo ideal es que la evaluación incluya objetividad y que capte la realidad, mediante procedimientos que eviten que las ideas preconcebidas e incluso los intereses personales, afecten los resultados de la evaluación. De la misma manera la información debe ser suficiente, aunque no necesariamente completa. La pretensión de exhaustividad suele ser, en muchas ocasiones, inconveniente y por lo tanto debe ser evitada. Se requiere un balance entre lo ideal y lo viable, considerando así si se justifica la mayor exhaustividad de la evaluación en función de los resultados a obtener. "Los métodos evaluativos presentan a menudo una transacción entre lo ideal y lo factible" :Weiss:89: 22:

A lo anterior debe agregársele también la validez y la confiabilidad, ya que ambos aspectos medirán a través de los instrumentos utilizados, lo que verdaderamente se quiere medir con calidad y estabilidad de la información, ésta pretende que los resultados no varíen con el evaluador, con la manera en que se efectúa la evaluación o con el instrumental utilizado.

Es interesante observar cómo ha ido evolucionando el campo de la evaluación, la investigación y la medición, cómo se ha proyectado el método convencional de evaluación en todos los ámbitos, con fuerte influencia, también en el ámbito de la administración institucional en todos sus campos de acción: producción, administración de recursos humanos, calidad, competitividad interna y externa, etc. Por lo tanto, la evaluación no debe ser concebida como una actividad aislada y autosuficiente. "Evaluar es fijar el valor de una cosa; para hacerlo se requiere de un procedimiento mediante el cual se compara aquello a evaluar respecto de un criterio o patrón determinado". :Franco:90:3:

Alternativamente, se ha definido a la evaluación como aquella rama de la ciencia que se ocupa del análisis de la eficiencia. También se ha definido a la evaluación como "el proceso encaminado a determinar sistemática y objetivamente la pertinencia, eficiencia e impacto de todas las actividades a la luz de sus objetivos. Se trata de un proceso organizativo para mejorar las actividades todavía en marcha y ayudar a la administración y planificación, programación y toma de decisiones futuras". :ONU :97:18:

Complementariamente, la evaluación ha sido caracterizada como "un medio sistemático de aprender empíricamente y de analizar las lecciones aprendidas para el mejoramiento de las actividades en curso y para el fomento de una planificación más satisfactoria mediante una selección rigurosa entre las distintas posibilidades de acción futura. Ello supone un análisis crítico de los diferentes aspectos del establecimiento y la ejecución de un programa, su pertinencia, su eficiencia y eficacia, su costo y su aceptabilidad para todas las partes interesadas" :OMS :97:11:

B. Conceptos Generales del Aspecto Estadístico

1. Orígenes de la Estadística:

La Estadística es una ciencia relativamente joven, ya que se presenta en una época menos remota. Dentro de sus principales componentes están: Blaise Pascal, Fermat, Laplace, Poisson, Bernoulli y Gauss. La creación de la primera cátedra de Estadística se produjo en Alemania durante el siglo XVII y su origen se atribuye a tres núcleos que hacen su apareamiento casi en forma simultánea.

- a) La Escuela Administrativa: Alemana, que considera los problemas del Estado. Sus principales representantes son:
- Vito Seckendorf (1626-1689)
 - Herman Conring (1600-1689)
 - Godofredo de Achenwall (1772-?)
- b) La Escuela Probabilística: Originalmente Italiana, pero primordialmente Francesa, que estudia los problemas relacionados con los juegos de azar.
- c) La Escuela Demográfica: Inglesa, preocupada por problemas actuariales, impulsada por Petty, Halley, King, Davenant y John Graunt.

Posteriormente y gracias a los trabajos de franceses como Borel, Frechet, Levy y los rusos como Tchevichev, Tchuprov, Markov, Kintchine y Kolmogrov, continuó el desarrollo de la escuela Probabilística, llegando a consagrarse en este siglo como Disciplina Matemática, el Cálculo de Probabilidades, base de la Estadística Inferencial.

Sin embargo, sólo hasta principios del siglo XX, una segunda escuela Inglesa, principalmente preocupada por problemas de investigación en Agronomía y Biometría, inició realmente y puso los cimientos de la Estadística como ciencia. Sus principales componentes son: Galton, Karl Pearson, "Student", R. A. Fisher.

Finalmente puede hablarse de una Escuela Escandinava, formada por actuarios, probabilistas y estadísticos tales como: Thiele, Charlier, Crámer, Gram.

De la Escuela Norteamericana se destaca Hotelling, Wald, Neymann, Wilks, Hoel, Mood, etc.; en la Escuela Hindú, cabe destacar la figura de Mahalanobis.

La Estadística, según afirman varios autores, proviene de tres raíces: del griego (statera) que significa "balanza"; del latín (status) que se refiere a "situación" y

del alemán (staat) que significa "estado". Cuando se escribe la palabra Estadística (con mayúscula) ésta se refiere a la ciencia y es la que se ocupa del análisis, descripción y predicción de fenómenos colectivos, medidos a través de un grupo de datos cuantitativos. Por otro lado, cuando escribimos la palabra estadística (con minúscula), nos referimos a colecciones sistemáticas de datos, relacionados a un fenómeno cualquiera que se ha medido.

Muchos autores, a través de la evolución de la Estadística, han dejado bien definidos algunos aspectos importantes en esta disciplina, con el propósito de distinguir con claridad y precisión, todos y cada uno de los conceptos que más adelante se describen. El nombre que le corresponde al profesional que se dedica a la Estadística o a las estadísticas es "Estadístico", mientras que estadígrafo se utiliza para designar cualquier función de datos empíricos utilizados con fines descriptivos. En resumen la Estadística como tal se configura de 1600 a la fecha y el proceso evolutivo de la misma se reafirma con las escuelas mencionadas. Su inicio coincide con la formación de algunas corrientes de la Matemática pura, especialmente la teoría del Cálculo de Probabilidades. La siguiente gráfica resume las corrientes que configuraron a la Estadística como disciplina científica.

EVOLUCION Y CONFIGURACION HISTORICA DE LA ESTADISTICA

MATEMATICOS Y ESTADISTICOS PRINCIPALES

| | | | | |
|----------------------------------------------------------------------------------------------------------------------------|--------------------|---------------------|-------------------|------------------|
| Vito Seckendorf (1626-1689) Herman Conring (1600-1689) Godofredo de Achenwall (1772) | 1 | | | |
| Mahalanobis | | | | 2 |
| Petty, Halley, King, Davenant, John Graunt (1620-1647) | 3 | | | 4 |
| Se inician y ponen cimientos de la estadística como ciencia, Galton, K. Pearson, "Student", R.A. Fisher | | | | 4 |
| Hotelling, Wald, Netman, Wilks, Hoel, Mood | | | | 5 |
| Thiele, Charlier, Crámer, Gram (actuarios, probabilistas y estadísticos) | | | | 6 |
| Blaise Pascal (1623-1647), Fermat (1601-1655), Laplace (1749-1827), Poisson (1777-1855), Bernoulli y Gauss. | 7 | | 8 | |
| definen la estadística probabilística: Borel, Frechet, Levy, Tchevichev, Tchuprov, Markov, Kintchine, Kolmocrov, Bernoulli | | | 9 | |
| 1600 | Siglo XVII 1700 | Siglo XVIII 1800 | Siglo XIX 1900 | Siglo XX 2000 |

Referencia de Escuelas: 1=Alemana, 2=Hindú, 3=Inglesa, 4=Inglesa, 5=Norteamericana, 6=Escandinava, 7=Francesa, 8=Rusa, 9=Italiana

2. Síntesis histórica de las estadísticas:

Las estadísticas son tan antiguas que no se tiene noción exacta o precisa de su origen; únicamente pueden ligarse a la configuración misma de las sociedades humanas. Desde su existencia se han producido censos, recuentos poblacionales, catastros, etc. referentes a información de índole económica, de recursos humanos. Se sabe que el año 2,000 A. De J.C. ya se levantaban en China, registros de esta naturaleza.

En el mundo de habla hispánica se tiene noticia de estadísticas obtenidas por los árabes en la península Ibérica, allá por el año 727 D. De J.C. Durante la época de los reyes Católicos y de Felipe II, etc., se levantaron censos bastante completos. Merece especial atención el famoso Censo del Marqués de la Ensenada (1748) que contiene variada información correspondiente a la España de Fernando VI.

El archivo de Indias (Sevilla) conserva valiosas estadísticas económicas que aún son consultadas por investigadores de diversa índole.

En términos muy generales, la historia mundial registra un constante y progresivo aumento de una inmensa, detallada y variada gama de estadísticas obtenidas por todos los países del mundo. Hoy día, los organismos nacionales e internacionales, públicos y privados en todo el mundo, producen constantemente innumerable cantidad de estadísticas de todo tipo.

3. Algunas Definiciones de Estadística y Métodos Estadísticos:

La cantidad de definiciones de Estadística existente hoy día, son contables; dentro de todas ellas las definiciones más utilizadas sobre Estadística, consideradas como clásicas, son enunciadas a continuación:

Estadística:

- a) "Estudio de los hechos morales o físicos del mundo que se prestan a enumeración o recuento y a comparación de las cifras a ellos referentes".

- b) Conjunto de métodos para efectuar decisiones adecuadas frente a la incertidumbre (Wallis y Roberts, *Statistics new approach*).
- c) Ciencia de tomar decisiones a base de las observaciones. (Clay Sprowls, *Elementary Statistics*).
- d) Operación de análisis matemático, que permite estudiar con el máximo de precisión los fenómenos incompletamente conocidos (J. Mothes, *Estadística Aplicada a Ingeniería*).
- e) Ciencia que trata de problemas relativos a las características operatorias de las reglas de comportamiento inductivo, basándose en experimentos aleatorios (Jerzy Neyman).
- f) Ciencia que trata de la recolección, organización, análisis e interpretación de los datos numéricos (IASI) (Instituto Interamericano de Estadística), de donde:
- “Recolección de datos” es el proceso de obtener medidas o recuentos. Sólo se obtienen conclusiones válidas con datos recolectados adecuadamente o que sean representativos.
 - “Organización de datos” es la tarea de presentar las medidas recolectadas en una forma adecuada para obtener conclusiones lógicas.
 - “Análisis de datos” es el proceso de extraer de las mediciones dadas, una información relevante con la cual se puede formular una descripción numérica, sumaria y comprensible.
 - Interpretación de los datos”, es la tarea de obtener conclusiones mediante el análisis de los datos; generalmente involucra la formulación de predicciones relativas a grandes conjuntos de elementos; utilizando la información disponible en una pequeña colección de elementos similares.

4. Métodos Estadísticos:

“Los Métodos Estadísticos son un instrumento en investigación y sirven para describir datos, estudiar relaciones causales, llegar a deducciones o inducciones en relación a conjuntos voluminosos de elementos o en algunos casos basándose en conjuntos menores de elementos o sub conjuntos de los mismos de donde provienen dichos datos” (G.V. Glass y J.C. Stanley).

- a) La Estadística comprende el conjunto de métodos y procedimientos para obtener, describir e interpretar conjuntos de datos y para basar decisiones y predecir fenómenos que pueden expresarse en forma cuantitativa. (ISI) (Instituto Internacional de Estadística).
- b) Los Métodos Estadísticos no se reducen ya sólo al análisis de datos demográficos, económicos y sociológicos, sino que se han extendido al campo de aplicación de todas las investigaciones en las que gran cantidad y complicación de los factores de variación, exigen una técnica de interpretación basada en el conocimiento de leyes. (A. Vessereau).

5. Tipos de Estadísticas:

Según la finalidad, uso y cobertura, las estadísticas pueden clasificarse básicamente en:

- a) estadísticas continuas: Todas aquellas que elaboran información por períodos de tiempo ininterrumpidos (anuales, quinquenales, bienales, semestrales, trimestrales, mensuales, quincenales, semanales, etc.).
- b) estadísticas discontinuas o esporádicas: Aquellas que elaboran información sobre fenómenos poco frecuentes (investigaciones específicas sobre fenómenos de interés muy particular o especial).
- c) estadísticas globales: Aquellas que proporcionan información sobre rubros muy generales sin mayor desglose.
- d) estadísticas de alto grado de detalle: Son estadísticas generalmente no publicables, existentes en archivos cuya característica es su detalle geográfico a nivel de divisiones políticas muy pequeñas.

- e) estadísticas para comparabilidad internacional: Son estadísticas ajustadas a presentación y nomenclatura especial, generalmente ajustada y proveniente de congresos, seminarios, reuniones.

6. Conceptos de "procesos" utilizados en el análisis estadístico de este modelo:

Cuando se habla de "procesos" para el análisis estadístico, éstos se refieren al desarrollo sucesivo de un fenómeno es decir, el seguimiento de un método sistemático para llegar a lograr un fin preciso, en el caso de los procesos estadísticos se refiere a un conjunto de operaciones a las que son sometidos los datos para conseguir un resultado final.

Por ejemplo, en la empresa de investigación, después de haber recopilado toda la información a través de las encuestas, dio inicio al primer paso para el proceso estadístico, que consistió en el arreglo ordenado de la información obtenida, es decir, la elaboración de un listado de los valores correspondientes a la muestra, de mayor a menor. Seguidamente se clasificaron los datos, de acuerdo a las reglas establecidas para las variables que más le interesaban a la empresa de investigación y esta clasificación de la información estadística generó las tablas con su respectiva información ordenada y simplificada. Terminado este proceso fue posible hacer cálculos adicionales, con medidas de tendencia central, variabilidad, correlación y regresión lineal.

a) Distribución de Frecuencias

Un conjunto de observaciones puede hacerse más comprensible y más significativo por medio de un arreglo ordenado, es más útil y práctico el resumen que se obtiene mediante la agrupación de datos. Antes de la era de las computadoras, uno de los principales objetivos de la agrupación de datos masivos era el facilitar el cálculo de varias medidas descriptivas como porcentajes y promedios. Debido a que las computadoras pueden ejecutar esos cálculos sobre grandes conjuntos de datos, sin agrupación previa, actualmente el propósito principal de ellos es resumir la información.

b) Correlación

Además de las medidas de promedios y tendencia central, variabilidad, asimetría, kurtosis y otras, hay necesidad de que se haga un análisis que describa el grado de relación entre dos o más variables. El coeficiente de correlación se calcula para conocer el grado de relación o asociación entre dos o más variables. Este tipo de análisis puede llevar a hacer predicciones sobre dicha relación de variables, es muy útil y usual en el análisis bidimensional y pluridimensional (muchas variables).

c) Análisis de Varianza

Se define como una técnica en la que la varianza total de un conjunto de datos se divide en varios componentes y cada uno de ellos se asocia a una fuente específica de variación, de manera que mediante el análisis es posible encontrar la magnitud con la que contribuye cada una de esas fuentes en la variación total. Los cálculos que se requieren para este análisis tiene un proceso largo y complejo por lo que la computadora asume una función importante y simplifica considerablemente el trabajo.

d) Variable

Una característica, cualidad o medición se considera como variable, si tal y como se observa, ésta toma diferentes valores en el tiempo. Las variables se aplican y atribuyen a fenómenos, entes, o cosas que son medibles en el espacio y en el tiempo. Las hay cuantitativas de muchas clases, pudiendo ser continuas o discretas. Las mediciones hechas sobre este tipo de variables, conllevan información respecto de cantidad, calidad o cualidad. Son cualitativas o categóricas: cuando se refieren a una característica difícil de cuantificarse, generalmente cualidad, contienen información respecto de los atributos.

7. Campos de la Aplicación de la Estadística:

Las actividades de toda persona o institución se desprenden de decisiones basadas en antecedentes de múltiple índole los cuales se recopilan o indagan con base en información de expresión o notación estrictamente cuantitativa. Tan es así, que el

hombre de negocios utiliza las estadísticas para determinar la reacción del público; el astrónomo confecciona tablas numéricas de la posición actual y futura de los astros; el auditor de una empresa no puede examinar todos los elementos que esta posee, pero mediante métodos adecuados selecciona muestras de los inventarios, acreedores y documentos de cobro; la semejanza entre los resultados observados y esperados en una experiencia genética se determina estadísticamente; el ingeniero industrial, que no puede analizar cada ampolleta fabricada, controla su calidad con base en muestras tomadas al azar (control de calidad); el actuario determina las primas de seguros de vida, jubilaciones, etc. mediante tablas estadísticas configuradas a propósito. El economista utiliza una amplia gama de estadísticas para los planes de consumidores y poder efectuar así, pronósticos de la tendencia de las actividades económicas. El gerente de una empresa eléctrica, proporciona un buen servicio a la comunidad, mediante la inspección de las variaciones estacionales de las necesidades de carga; el sociólogo trata de auscultar la opinión pública, mediante encuestas para determinar la preferencia por un candidato presidencial X o su oposición frente a determinados problemas políticos y sociales; los investigadores agropecuarios determinan la significación de ensayos agrícolas mediante procesos estadísticos; el geólogo aplica los métodos estadísticos para saber la edad de las rocas, el filólogo emplea el análisis cuantitativo para determinar la época de separación de dialectos y lenguas de origen común; el crítico literario considera la frecuencia del uso de diversas partes de la oración para estudiar el estilo de diferentes autores, el educador evalúa el rendimiento de sus educandos por medio de diversos métodos estadísticos, el planificador multisectorial o unisectorial necesita de estadísticas para planear sus proyectos; el investigador en general para diagnosticar cualquier fenómeno, requiere sondear el problema que le preocupa y para ello recurre tanto a las estadísticas como a la Estadística.

a) Estadística en el Campo Social

La estadística viene a ser una poderosa herramienta para aumentar el conocimiento y se ha desarrollado fuertemente en los últimos años. Sus aplicaciones son bastante amplias especialmente en las Ciencias Sociales. La Ciencia Política hace sondeos para conocer las preferencias de un candidato sobre otro. Los psicólogos se apoyan en la Estadística para hacer escalas de medición de pruebas y de pacientes; los fabricantes de medicamentos basan sus

dosificaciones en los resultantes que obtengan después de hacer bastantes muestreos y aplicación de métodos estadísticos. Por su utilidad, podemos ver que en cualquier ciencia o rama, la estadística juega un papel muy importante.

Uno de los principales objetivos de la estadística es ayudar en la evaluación científica de ciertas afirmaciones. La mayoría de personas, empieza a convencerse de que la comprensión de la estadística tiene importantes aspectos prácticos para el éxito en su vida. Actualmente se lleva una vida muy agitada y ésta es bombardeada por "autoridades" que dicen, con base en "afirmaciones ciertas", lo que hay que hacer, cómo se debe vivir, lo que se debe comprar, lo que se debe valorar, etc.

En resumen, las disciplinas en las que la Estadística se aplica con más énfasis son: ciencias físico-matemáticas, biología, meteorología, ciencias agrícolas, psicología, sociología, análisis de mercado, control de fabricación, educación, medicina, ingeniería, veterinaria, economía, arqueología, ecología, historia, geología, demografía, filología, moneda, banca y física, etc.

b) Sistemas Computarizados y Estadística:

Como se menciona al principio de este trabajo profesional, el adelanto tecnológico se deja ver en casi todas las áreas que sirven de apoyo al ser humano para realizar sus tareas. Cada vez es más significativo tener la herramienta adecuada para lograr una mayor eficiencia en el desempeño diario. El uso de computadoras en el área de la estadística, se incrementó en gran medida durante los últimos diez años.

En actualidad, la mayor parte de datos que son analizados para algún tipo de investigación, utiliza programas estadísticos en computadora, en lugar de tener que hacer el mismo análisis, utilizando una calculadora y lápiz. Las computadoras tienen muchas ventajas, pues ahorran tiempo, esfuerzo y minimizan la posibilidad de error de cálculo por la intervención humana, especialmente cuando se manejan datos masivos, pero debe saberse utilizar con eficiencia, de lo contrario estamos derrochando tiempo, costo y tecnología.

Entre los programas de computadora que ayudan al análisis estadístico, los más populares son el Statistical Package for the Social Sciences (SPSS), (Paquete Estadístico para las Ciencias Sociales), Biomedical Computer Programs-P series (BMDP), (Series de Programas de Computadora Biomédicos), Statistical Analysis Systems (SAS), (Sistemas Analíticos Estadísticos), SYSTAT, (Sistemas Estadísticos) y MINITAB (Software para Estadística, especialmente utilizado para didáctica). Estas versiones de programas para apoyo estadístico, vienen en diferentes formatos, que pueden utilizarse en computadoras personales bajo ambientes de sistemas operativos como Windows, Linux, DOS, y también en formatos para computadoras de tipo Macintosh o computadoras grandes llamadas mainframes y las intermedias que se les conoce como minicomputadoras.

“Cuando se comienza a resolver problemas con MINITAB y SPSS, es un inicio directo a experimentar la diversión y la capacidad que puede otorgar la computadora en su estudio de la estadística. De hecho, una vez que se haya utilizado una computadora para analizar datos, es probable que se pregunte “¿por qué tengo que hacer todos estos cálculos complejos a mano?” Desafortunadamente, el uso de una computadora para calcular el valor estadístico no le ayuda a comprenderlo. El entendimiento del estadístico y su mejor uso se logran a través de la realización de cálculos “a mano”, principalmente, y al emplear las computadoras como un medio de apoyo. La computadora es particularmente eficaz para promover la comprensión, al utilizarse en presentaciones gráficas. Por supuesto, una vez que se aprenda todo lo que pueda a partir de los cálculos “a mano”, el uso de la computadora para obtener valores estadísticos parece bastante razonable”. Pero sobre todo porque se tiene la capacidad de poder analizar e interpretar esos resultados que se obtuvieron.

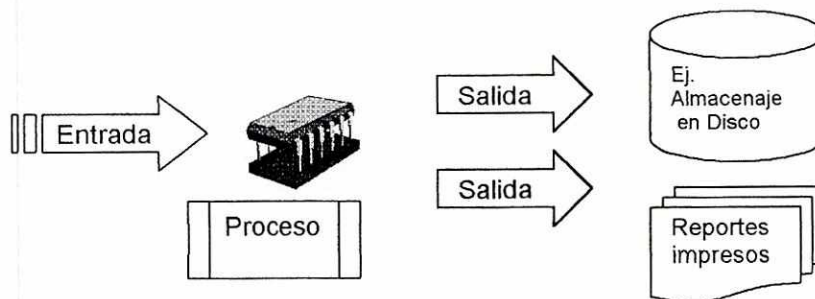
C. Conceptos del Sistema de Cómputo

1. El Sistema Operativo

Antes de llegar a las definiciones de Sistema Operativo, es pertinente citar algunos conceptos respecto de un computador, considerándolo como un conjunto de componentes electrónicos en donde se realizan 4 funciones básicas relacionadas con la información, las cuales son: **la entrada, el proceso, la salida y el almacenamiento.**

Uno de esos componentes electrónicos será un componente electrónico principal que se encarga de hacer el cómputo de información necesaria. También habrá una memoria interna que en algunos ámbitos de lógica se le llama RAM (Memoria de Acceso Random). En algunos otros ambientes de trabajo también se le conoce como central. Tanto el procesador como la memoria están físicamente residentes en un circuito impreso de distribución, que también se le conoce como Circuito Impreso Principal o Tarjeta Madre.

En la descripción anatómica puede decirse que está en el corazón del computador, es en donde se encuentran los componentes electrónicos y programas lógicos controladores de los demás miembros del cuerpo completo con una función específica. Para que se efectúen las cuatro funciones anteriormente mencionadas, se necesitarán subsistemas que permitan la interacción humana con electrónica, tales como un monitor o pantalla, un teclado y unidades de almacenamiento magnético de datos. Todos estos subsistemas están conectados a la Tarjeta Madre, por medio de sus controladores electrónicos, quienes ayudarán al procesador a ejecutar sus procesos de I/O (Input / Output) (entradas - salidas). Se ilustra gráficamente de la siguiente manera:



Todas las partes internas del computador están conectadas por un bus. Físicamente el bus es el componente electrónico, donde hay una conexión con los controladores. Grant Taylor, en su libro Linux, describe el bus como "la autopista de datos entre el procesador, su pantalla, su disco y cualquier otro componente". :Taylor: 43:

El procesador, que es el que tiene control sobre todo, realmente no puede ver cualquiera de los sub sistemas directamente; tiene que comunicarse con ellos por medio del bus. El "único" sub sistema con el que tiene acceso inmediato, es la memoria central. A fin de que los programas funcionen, deben tener representación en la memoria. Se dice: "deben estar en memoria"

2. Anatomía Básica de un Sistema Operativo

Anteriormente se hizo una descripción de cómo se integra un conjunto de componentes electrónicos para manejar información. Sin embargo si un computador no tiene un programa o un conjunto de instrucciones que siguen un orden lógico, será una magnífica pieza de electrónica inerte, incapaz de cumplir con las cuatro funciones.

Cuando se inyecta energía al computador o se dice que se enciende, la primera cosa que éste tiene que hacer, es empezar una serie de instrucciones o programas llamados Sistema Operativo. El Dr. Chin menciona en su libro "Disk Operating System" (Sistema Operativo en Disco), que este sistema "es un conjunto de programas que supervisan las operaciones del computado". :Chin:92:130:

Puede resumirse que el trabajo del Sistema Operativo, es ayudar a otros programas del computador, a que funcionen apropiadamente, manejando los detalles complejos que se requieren para controlar todo el hardware¹ de la computadora.

Ese proceso de arranque, que se le denomina bootstrapping (carga inicial), será el punto de partida para fusionar todos los componentes electrónicos con el Sistema Operativo y convertir el computador en un sistema activo y funcional.

¹ Ver glosario, pág. 92

3. Una Red de Computadoras

Cuando se habla de una red de computadoras, se refiere a un grupo de varias computadoras, con un componente electrónico especial, que se eslabonan entre sí por medio de un cableado especial. Al estar conectadas entre ellas, pueden ser compartidas en ambas, diferente software² de aplicación, es decir un conjunto de programas con una tarea específica que le interesa a un usuario; asimismo se puede compartir archivos para uso múltiple, programas de apoyo o herramientas de usuario y también la mayor parte de recursos electrónicos o dispositivos, que generalmente están en los computadores, tales como impresoras, discos duros, unidades de respaldo, unidades de CDs, etc.

En pocas palabras, se pueden compartir los recursos del computador, con otros usuarios que estén conectados. Esta conexión puede ser de tipo local o remota. En el caso de local, se le conoce, por sus siglas en Inglés, como un LAN (Local Area Network), que quiere decir Red de Area Local. El otro tipo de red, que no es de conexión local, sino de acceso remoto, se le conoce por sus siglas en Inglés RAN (Remote Area Network), que quiere decir Red de Area Remota. A continuación puede verse la descripción de ambas.

a) Red de Area Local

Este tipo de red tiene varias computadoras que comparten recursos entre sí, siempre y cuando estén conectadas entre ellas por medio de algún dispositivo físico como cables, conectores y concentradores³ de distribución. En este tipo de Red de Area Local, las topologías⁴ que más se han diseñado son: la de anillo, la de bus y la de estrella.

²/ Ver glosario, pág. 95

³/ Ver glosario, pág. 89

⁴/ Ver glosario, pág. 95

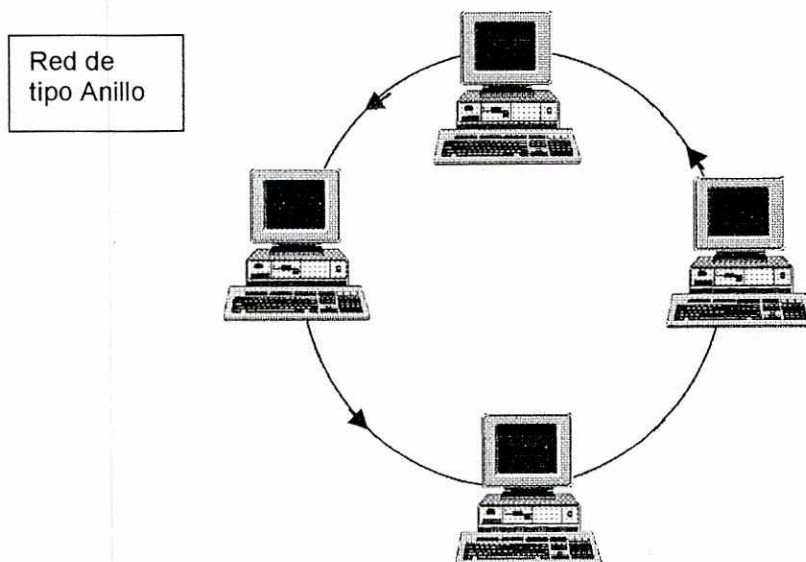
b) Red de Tipo Anillo

Este tipo de interconexión requiere una conexión de punto a punto entre cada uno de los nodos (puntos de conexión) de la red. La vía de datos de la red debe formar un anillo cerrado, que permite que los datos puedan viajar en una dirección y retornar al punto de conexión que lo envió originalmente. El cableado de punto a punto de este tipo de red, requiere que en cada nodo haya una señal de una repetidora. La repetidora, debe recibir una señal de entrada y regenerar la señal de salida para el siguiente nodo.

Este diseño de anillo puede convertirse en un diseño que si no se toman las medidas de respaldo, ofrecerá las deficiencias en velocidad y especialmente que si algún nodo no está activo o tiene alguna falla de transmisión, la vía de datos se ve interrumpida para el resto de usuarios que estén conectados en la red.

Este diseño de topología fue utilizado, por mucho tiempo, en sistemas exclusivos de IBM, con computadoras personales que tenían una tecnología de bus microcanal, propio de la misma corporación.

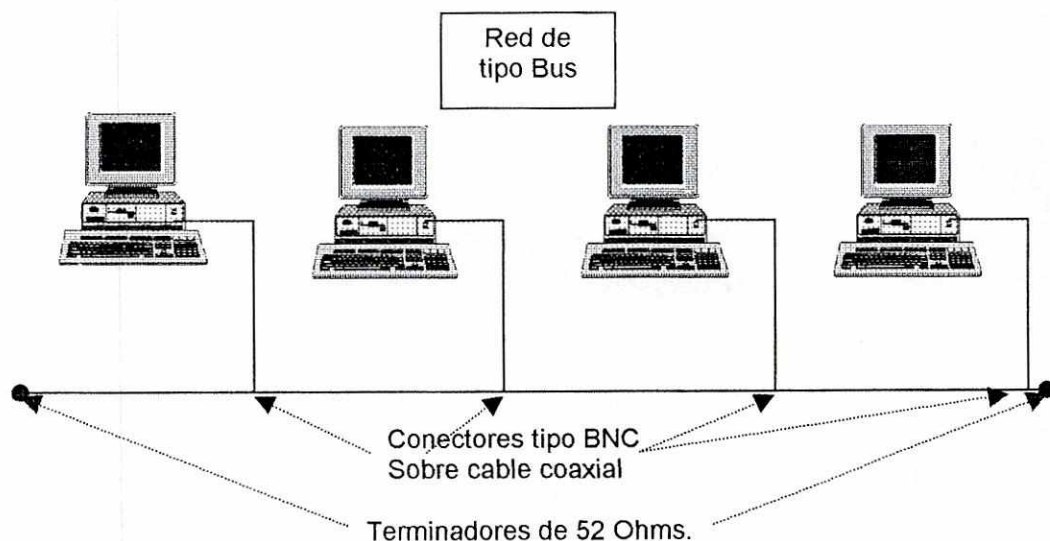
La diversidad de equipo electrónico que empezó a proporcionar alternativas, en lo que era arquitectura abierta, es decir un común denominador para el intercambio de dispositivos electrónicos, sin importar la marca, abrió las puertas a nuevas topologías que dieron evidencias de mayor eficiencia al usuario final. Puede verse a continuación un esquema que ilustra este tipo de topología.



c) Red de tipo Bus

Esta topología fue utilizada desde los comienzos de las redes de área local y aún es común en nuestro medio. Como su nombre lo indica, esta red provee un bus común para todos los puntos de conexión de la red y sus dispositivos. Los puntos de conexión transmiten datos a través de un cable físico, de tal manera que todos los nodos pueden ver cada uno de los paquetes de data viajando en él. No necesariamente hay una estación que controla esto en la red. Cada uno de los nodos⁵ debe incluir la inteligencia necesaria para colocar los datos en el bus y que puedan ser extraídos del mismo por el interesado. A esta inteligencia se le conoce como protocolo.

Esta topología ha estado dentro de las más utilizadas, por su versatilidad y facilidad de utilizar interfaces, no importando una marca específica, también porque en determinado momento un usuario puede desconectarse físicamente de la red o del bus mismo, sin que afecte al resto de usuarios. Sin embargo, por utilizar en su mayor parte cable de tipo coaxial, no va a ser el de mayor velocidad, además de que causó muchos problemas, porque en repetidas situaciones si los conectores de tipo BNC eran mal ensamblados por los técnicos, podía interrumpirse el flujo de datos y no se sabía inmediatamente, dónde estaba el defecto. Vea un esquema de este tipo de red.

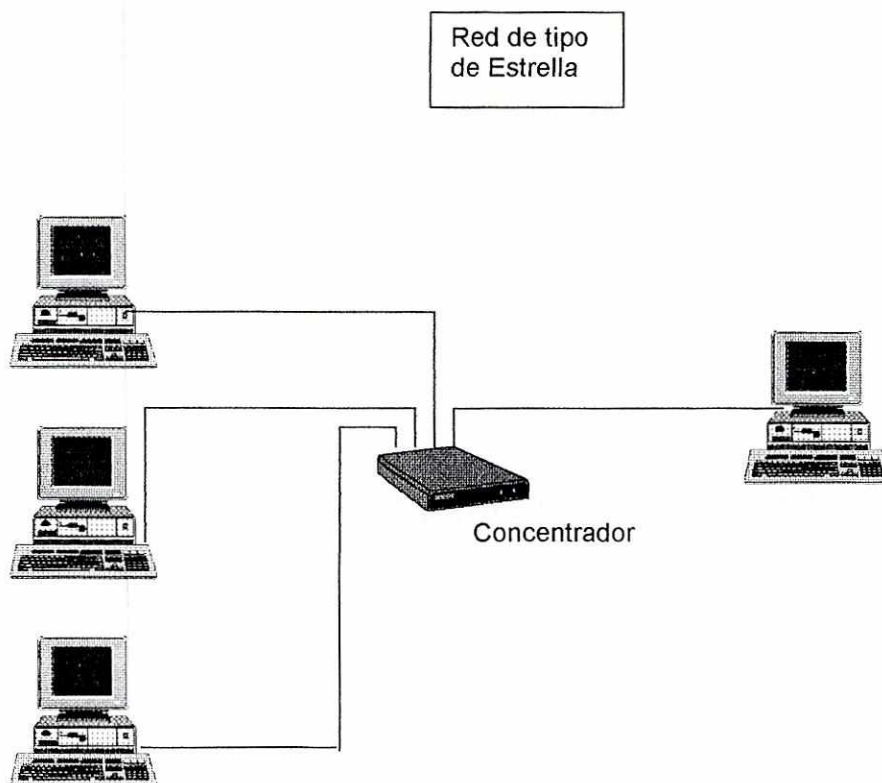


⁵/ Ver glosario, pág. 94

d) Red de tipo Estrella

La topología de estrella, para efectos de ejemplificación, es típica de un sistema telefónico, donde cada nodo se conecta con un concentrador⁶ central en una configuración de punto a punto. En el caso de computadores, estos serán los nodos al igual que un aparato de telecomunicaciones. El conector central es el responsable de conectar un nodo con otro, si es que los dos necesitan comunicarse. Este tipo de topología requiere mayor cantidad de cable que las otras topologías. También provee de la forma más simple para diagnosticar cuando ocurre algún error de comunicación entre nodos, ya que cada nodo tiene una y única conexión central.

El problema se aísla para diagnóstico y no afecta los otros nodos en la red, sin embargo si el concentrador central llega a fallar, toda la red falla también. A continuación se presenta el esquema que ilustra este tipo de topología.



⁶/ Ver glosario, pág. 89

e) Red de Area Remota

Al igual que la red local, tiene como propósito compartir los recursos de hardware como de software entre diferentes usuarios. Este tipo de red cumple con este cometido, para usuarios que no necesariamente están conectados entre sí por medio de un cable físico. Como su nombre lo indica, el acceso a los diferentes recursos será por medio de una comunicación remota, ya sea por líneas de comunicación telefónica o por ondas de radio frecuencia.

Los protocolos⁷ de comunicación, para que se entiendan entre computadores a distancia, deberán poseer capacidades especiales de red, para corrección de errores automáticamente, asimismo deben poseer un robusto soporte a problemas de determinación en el envío y recibo de datos empacados, como el ruteo de datos a las direcciones necesarias, tanto del computador que envía como del que recibe.

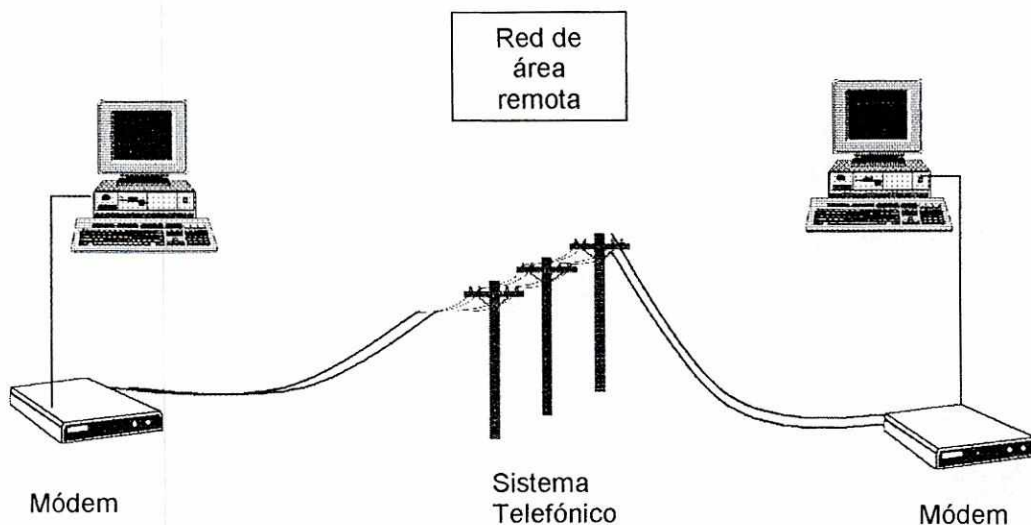
Estas redes permiten a usuarios múltiples, acceder un recurso simultáneamente a través de un dispositivo conocido como módem. Un módem se requiere que esté en cada punto de enlace, es decir el local y el remoto de quienes desean conectarse. El módem es el dispositivo electrónico que debe ser instalado entre el sistema de cómputo y el sistema telefónico. Este dispositivo es el eslabón final entre el sistema digital y un sistema análogo de comunicaciones. Si no fuera por el módem, el mundo de los negocios de comunicaciones de datos no sería posible. Asimismo, en el caso de los programas aplicados, no sería posible trabajarlos a distancia.

El nombre módem⁸, viene de las funciones **MOD**ulación-**DEM**odulación. Con lo cual decimos que el módem recibe la señal binaria de una terminal o computador y la convierte en señales de frecuencia de voz o señales de tono. De allí, el módem transmite estos sonidos a través del sistema de telefonía pública. En la otra punta hay otro módem compatible al primero, el cual convierte todos estos sonidos en señales eléctricas binarias, las que a su vez son enviadas a la terminal o computador. En algunos casos especiales, el módem puede enviar estas señales directas a una

⁷/ Ver glosario, pág. 95

⁸/ Ver glosario, pág. 93

impresora de tipo de comunicación serial. A continuación, se puede ver de forma gráfica la representación de una red remota.



f) Plataformas de Operaciones de una Red

En la sección "Anatomía de un Sistema Operativo", se menciona que: para que tome vida un sistema de computación, se necesita no solamente el hardware o el conjunto de componentes electrónicos, sino también el conjunto de programas que componen el Sistema Operativo. Este Sistema Operativo es la plataforma sobre la cual están edificados los sistemas que operan independientemente, sin ninguna red y también los que están en un ambiente de compartir recursos entre usuarios. A esta plataforma, se suman los programas de enlace o de comunicaciones para la red, los cuales pueden ser para comunicación local o remota como se ha descrito en los párrafos anteriores.

Dependiendo de las diferentes necesidades de cada usuario, empresa y/o institución, así será escogido el computador como pieza de hardware, con todos sus dispositivos y simultáneamente el Sistema Operativo adecuado para proveer la capacidad de almacenaje, como la velocidad de procesos y transmisiones de datos. Independientemente de la topología que se emplee, es necesario definir si se establecerá una clase de red de punto a punto, donde prácticamente todos pueden

compartir recursos de hardware como de archivos y programas o si se estará utilizando un computador como servidor dedicado, exclusivamente para emplearse de almacenamiento masivo centralizado.

La segunda opción provee de mejores alternativas para velocidad y seguridad, sin embargo puede llegar a ser la opción más elevada en costos, por tener un computador que no está libre para operaciones por un determinado usuario. La opción de punto a punto ofrece la ventaja que no tiene que haber un computador específico, que tiene que actuar como servidor y que si éste no está habilitado, los demás usuarios no pueden trabajar. A pesar de ello, en la administración de sistemas, se puede nombrar a un computador, como el máster donde hay ciertos archivos de interés para todos o bien donde reside un programa específico para cumplir una tarea, independientemente de los programas o funciones específicas que estén sucediendo en cada uno de los nodos conectados a la red.

En algunos casos, los programas adicionales para comunicaciones de la red, vienen como parte de módulos de integración del Sistema Operativo, sin embargo en algunas ocasiones se necesita instalar programas adicionales de otros fabricantes que se dedican a esta especialidad. Por ejemplo, un caso típico que se da en nuestro ambiente: Se desea instalar un servidor⁹ dedicado, el cual contiene un programa de facturación y cuenta corriente para ser compartidos con 9 usuarios en total. Algunos están facturando en el departamento de ventas y otros recibiendo pagos por facturas emitidas. Después de considerar muchos factores, se decide instalar un sistema de software de enlace de red elaborado por la casa Novell, sin embargo su plataforma original para arranque del servidor y de las estaciones conectadas, es DOS, como Sistema Operativo.

Por otro lado hay un usuario que desea correr las herramientas hechas para trabajo de oficina en ambiente de Windows 98, en su computador personal, pero a la vez necesita utilizar el mismo programa de facturación y cuentas corrientes.

Muy bien; se puede utilizar el mismo ambiente que se describe en el ejemplo anterior, pero ciertas computadoras tendrían que tener un arranque dirigido por

⁹/ Ver glosario, pág. 95

Windows 98, el cual utiliza la plataforma de DOS, sin embargo los programas para comunicarse con el resto de la red, ya vienen integrados en Windows 98 y éstos no necesitan otro módulo separado de un determinado fabricante.

Otro caso, que se da muy a menudo en oficinas pequeñas: es cuando un usuario solamente quiere utilizar las herramientas que provee el integrador Office 97, pero desea compartir toda clase de recursos de software, entre tres usuarios.

Una solución práctica es seguir utilizando, su Sistema Operativo actual, que puede ser en este caso Windows 98, Windows 2000, Windows Millennium y utilizar los módulos de conexión que ya vienen dentro de esos paquetes, para tener una red de punto a punto, lo que en Inglés sería decir "Peer to Peer"

Esta opción proporciona al usuario lo que necesita, sin tener que invertir en un servidor dedicado o la instalación de software de red, los cuales elevarían el costo considerablemente y además se necesita de soporte técnico de alto nivel para administrar la red completa.

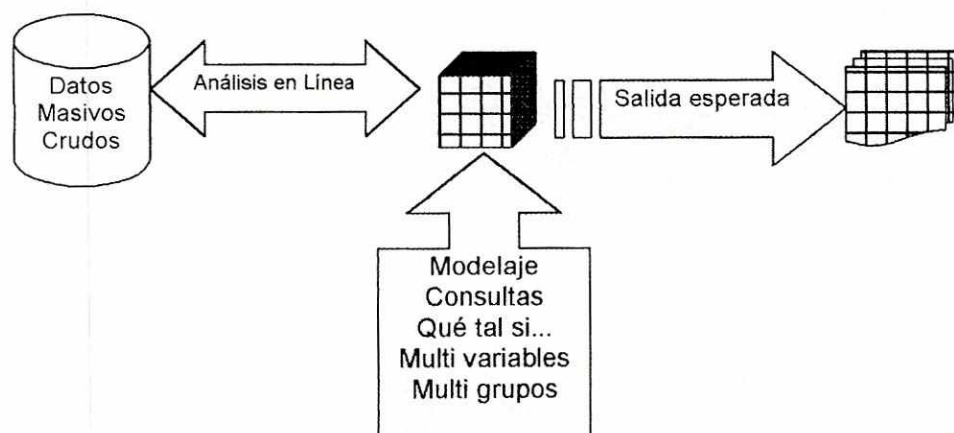
4. Generalidades de Olap

Por sus letras en Inglés: On Line Analytical Processing (Procesamiento Analítico en Línea) En todo el sentido de la palabra, cuando hablamos de OLAP estamos diciendo sencillamente, que es un procedimiento donde existe un análisis de datos, que está sucediendo en el momento mismo que tenemos nuestra base de datos abierta, con todas sus variables y además con todas sus clasificaciones de grupos y sub grupos. *"El cual incluye navegación entre datos, consultas, modelaje, pronósticos, que tal si... y data mining (minado de datos)"*.

Muchas personas no utilizan apropiadamente el término, pues a veces piensan que cualquier programa que produzca reportes multidimensionales, será OLAP, sin embargo podemos decir que ese producto tiene esas extensiones de reportes a las capacidades completas de la herramienta.

A esto se refiere Jeff Stamen, Senior VP de Oracle, diciendo que la metodología OLAP no se refiere únicamente a producir una salida o reporte de datos multi

dimensionales como son los reportes llamados de cubo, sino más específicamente se refiere al análisis en sí de la data cruda, a través de diferentes capas y profundidades, donde a veces se necesitará crear un porcentaje de data multi dimensional para entonces extraer un reporte con esas características. En relación a esto, menciona el anterior autor, existen varias herramientas para presentar reportes multidimensionales, pero a cambio de que haya una base de datos ya multi dimensionada. Esto no necesariamente es OLAP. El siguiente diagrama es una representación gráfica del concepto OLAP.



Un presidente de la compañía Oracle, refiriéndose al tema de OLAP, hizo el siguiente comentario: *"En cierta ocasión se discutía sobre la necesidad de que exista una base de datos con varias dimensiones, para poder utilizar herramientas de tipo OLAP y alguien dijo: Qué importa si es o no de múltiples dimensiones, lo que importa son los resultados; que sean presentados en una forma cúbica (tridimensional). A lo cual alguien más respondió: es cierto debe quedar en segundo plano pero no podemos dejar de pensar que sí tiene relevancia porque hay muchos riesgos en cargar, administrar y tener que ver con los datos, inclusive que todas las cosas estén lo más cerca o iguales a su estado original.*

El procesamiento analítico en línea u OLAP es una tecnología cada vez más popular que puede mejorar dramáticamente el análisis de negocios, pero también se ha caracterizado históricamente por herramientas costosas, implementaciones difíciles y distribución inflexible.0

OLAP es un componente clave en el proceso de almacenamiento de datos (*data warehousing = bodega de datos*) y estos servicios proporcionan la funcionalidad esencial para una gran variedad de aplicaciones que van desde reportes corporativos hasta soporte avanzado de decisiones. Por ejemplo la inclusión de la funcionalidad OLAP dentro del SQL Server ayuda a hacer el análisis multidimensional más accesible y trae los beneficios del OLAP a una audiencia mayor. Esta audiencia incluye no sólo a organizaciones más pequeñas sino a grupos e individuos dentro de grandes corporaciones que han sido excluidos de la industria del OLAP por el costo y la complejidad de los productos actuales.

Históricamente, el grueso de la inversión corporativa en cómputo ha sido en sistemas que generan o capturan información o datos, tales como contabilidad, manufactura o información de clientes. Cada vez más, las organizaciones invierten en aplicaciones y tecnologías que proporcionan un valor agregado de esta información recolectada. El almacenamiento de datos es el proceso de recolectar, limpiar y cernir datos o información de varios sistemas operacionales, y hacer disponible la información resultante a una amplia audiencia de usuarios finales para ser analizada o reportada. El almacenamiento de datos (*data warehousing*) y la tienda de datos (*data mart*) son utilizados para describir estos almacenes no volátiles de información limpiada y resumida que están disponibles para que la vean los usuarios.

A pesar del proceso de almacenamiento de datos de preparar información para el consumo del usuario final, la mayoría de la información en un almacén de datos relacional no se encuentra fácilmente. Generalmente, las estructuras de datos son difíciles de comprender por el usuario final, o las preguntas (tales como "¿Quiénes fueron los mejores vendedores de cada región durante el año pasado, mensualmente?") son complejas cuando se expresan en SQL. Algunos retos pueden ser enfrentados con herramientas avanzadas de peticiones (*queries*), las cuales ocultan al usuario final la complejidad de la base de datos. Como sea, para una gran clase de aplicaciones donde el usuario final está viendo información multidimensional, se cree que la solución óptima es la tecnología OLAP.

Todas las organizaciones tienen datos de varias dimensiones y la complejidad no es necesariamente una función del tamaño de la compañía. Aun a las más pequeñas compañías les gustaría poder rastrear sus ventas por producto, vendedor, geografía,

cliente y tiempo. Las organizaciones han buscado durante mucho tiempo herramientas para acceder, navegar y analizar información multidimensional de una manera fácil y natural.

OLAP proporciona a las organizaciones los medios para acceder, ver y analizar los datos con alto desempeño y flexibilidad. Primero y más importante, OLAP, presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de navegación, los usuarios finales pueden ver y entender más efectivamente la información de sus almacenes de datos, permitiendo así a las organizaciones reconocer mejor el valor de sus datos.

En segundo lugar, el OLAP acelera la entrega de información a los usuarios finales que ven estas estructuras multidimensionales al preparar algunos valores computados en los datos por adelantado, en vez de al momento de ejecutarse. La combinación de navegación fácil y rápido desempeño le permite a los usuarios ver y analizar información más ágil y eficientemente que lo que es posible solamente con tecnologías de bases de datos relacionales. El resultado final: se pasa más tiempo analizando los datos y menos tiempo las bases de datos.

Un principio clave del OLAP es que los usuarios deberían ver tiempos de respuesta consistentes para cada vista, o rebanada, de datos que requieran. Dado que la información se colecta en el nivel de detalle solamente, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados, o agregaciones, son la base de las ganancias de desempeño del OLAP.

En los primeros días de la tecnología OLAP, la mayoría de las compañías asumía que la única solución para una aplicación OLAP era un modelo de almacenamiento no relacional. Después, otras compañías descubrieron que a través del uso de estructuras de base de datos (esquemas de estrella y de copo de nieve), índices y el almacenamiento de agregados, se podrían utilizar sistemas de administración de bases de datos relacionales (RDBMS) para el OLAP. Estos vendedores llamaron a esta tecnología OLAP relacional (ROLAP). Las primeras compañías adoptaron entonces el término OLAP multidimensional (MOLAP).

Las implementaciones MOLAP normalmente se desempeñan mejor que la tecnología ROLAP, pero tienen problemas de escalamiento. Por otro lado, las implementaciones ROLAP son más escalables y son frecuentemente atractivas a los clientes debido a que aprovechan las inversiones en tecnologías de bases de datos relacionales preexistentes.

Un desarrollo reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalamiento. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

Un reto fundamental en la implementación del OLAP es mapear el esquema inicial de la base de datos a un modelo multidimensional. Esto requiere de un significativo esfuerzo de programación con muchos de los productos en el mercado hoy en día. En la evolución de los productos OLAP, el diseño de la base de datos OLAP se ha vuelto un proceso especializado y arcano, intrincadamente enlazado a la tecnología específica del OLAP que se esté implementando. Consecuentemente, los desarrolladores de bases de datos OLAP son muy especializados, lo cual ha llevado a altos costos de desarrollo de aplicaciones concentrados en la etapa de diseño de datos.

En la mayoría de las implementaciones de OLAP, se asume que los datos han sido preparados para el análisis a través del almacenamiento de datos (*data warehousing*) y que la información se ha extraído de sistemas operacionales, limpiado, validado y resumido antes de incorporarse en una aplicación OLAP. Este es un paso vital en el proceso, que asegura que los datos que son vistos por el usuario OLAP son correctos, consistentes y que llenan las definiciones organizacionales para los datos.

Cada vez más, la información en un almacén de datos se organiza en esquemas de estrella o de copo de nieve, lo cual simplifica el entendimiento de los datos por parte del usuario, maximiza el desempeño de las peticiones (*queries*) de la base de datos para aplicaciones de soporte de decisiones y requiere menor espacio de almacenamiento para bases de datos grandes.

En el caso de SPSS, presenta un acercamiento de OLAP cuando hace un análisis estadístico en línea. Lo cual significa que simultáneamente, cuando en una base de datos hay cambios de datos, automáticamente estos se pueden observar en las tablas de resultados, sin necesidad de ejecutar de nuevo los procedimientos o las pruebas estadísticas que inicialmente permitieron ver esos resultados. Ahora bien, más que una acción dinámica todo el tiempo que se tiene abierta la sesión, se puede combinar con los reportes de cubo y es allí donde se puede apreciar mejor, la ventaja de esta característica, pues se puede observar de inmediato los resultados no sólo en una dimensión pero en tres. Estos resultados pueden ser de tipo gráfico como de tipo numérico o de textos y permite utilizar simultáneamente, varios procedimientos estadísticos. Ejemplo: El total y promedio de las ventas para diferentes regiones y líneas de productos dentro de las regiones mismas. Se refiere entonces al procesamiento analítico interactivo, calcula totales, medias y otros estadísticos univariados para variables de resumen continuas dentro de los rangos de una o más variables categóricas de agrupación.

En resumen, algunas estadísticas que puede ofrecer son por ejemplo: número de casos, medias, medianas, medianas agrupadas, error estándar de la media, mínimos, máximos, rangos, valores de datos agrupados, valores clasificados por categoría de datos agrupados, desviación estándar, varianza, kurtosis, error estándar de kurtosis, sesgos, error estándar de sesgos, porcentaje del total de casos, porcentaje de la suma total, porcentaje del total de casos dentro de las variables agrupadas, porcentaje de la suma total dentro de las variables agrupadas, media geométrica y media armónica, sin embargo, para efectos de la presentación de este modelo, más adelante o en el glosario sólo se describen los conceptos de los procedimientos utilizados en el mismo.

a) Reportes Estadísticos de Cubo

En un modelo de datos OLAP, la información es vista como cubos, los cuales consisten de categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo de datos multidimensional simplifica a los usuarios el formular peticiones complejas, arreglar datos en un reporte, cambiar de datos de resumen a datos de detalle y filtrar o rebanar los datos en subconjuntos significativos. Por ejemplo, las dimensiones típicas de un cubo que contenga información de ventas, incluiría tiempo,

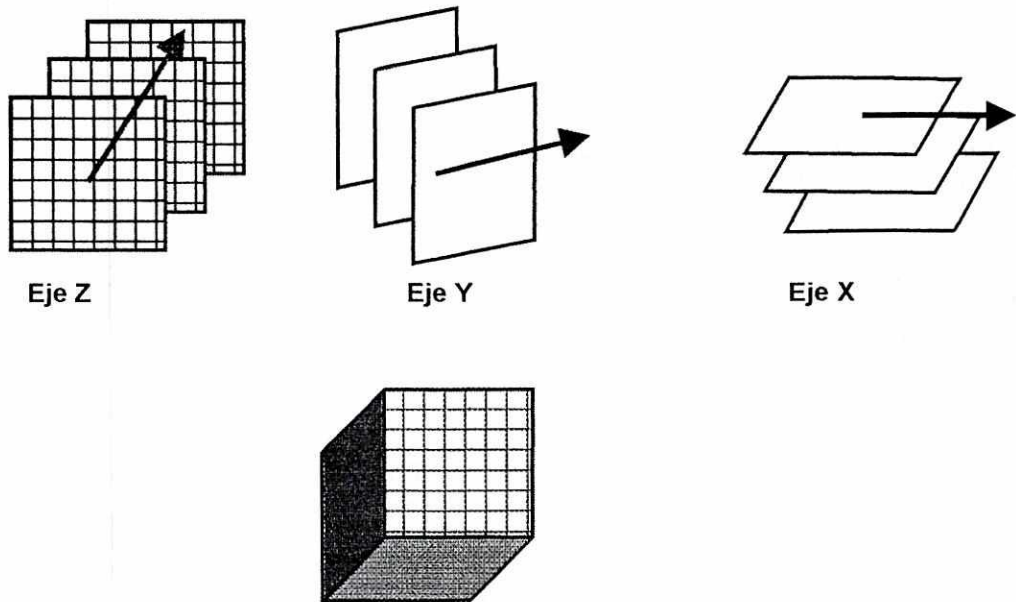
geografía, producto, canal, organización y escenario (planeado o real). Las medidas típicas incluirían ventas en dólares (u otra moneda), unidades vendidas, número de personas, ingresos y gastos.

¡Error! Marcador no definido. Dentro de cada dimensión de un modelo de datos OLAP, los datos se pueden organizar en una jerarquía que represente niveles de detalle de los datos. Por ejemplo, si una empresa quisiera evaluar su departamento de producción, dentro de la dimensión de tiempo y geografía, se puede tener estos niveles: años, meses y días; de manera similar, dentro de la dimensión geografía, se puede tener estos niveles: país, región, estado/provincia y ciudad. Una instancia particular del modelo de datos OLAP tendría valores para cada nivel en la jerarquía. Un usuario que vea datos OLAP se moverá entre estos niveles para ver información con mayor o menor detalle.

Los reportes de cubo, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos. Sin embargo, el solo hecho de describir el modelo de datos en una forma más intuitiva, hace muy poco para ayudar a entregar la información al usuario más rápidamente. La idea central que está detrás de los reportes de cubo, es que si se tiene acceso a los datos, de forma en línea, como lo permite OLAP, no importando el proceso que se esté desarrollando en ese momento, ¿por qué no poder producir un reporte al antojo del usuario final o del analista que estudia el comportamiento de datos en una simulación de modelo? Lo que hace interesante los reportes de cubo, es el hecho de que se puede generar un reporte de datos, desde varias dimensiones, como si se tuviera un cubo, con varias columnas y filas de varias profundidades. Se puede escoger la capa o la columna que se necesite presentar, rotarla sobre un eje antojadizo y extraer la información necesaria sin perder el análisis estadístico que se está procesando en el momento.

Se puede girar sobre un eje, la bandeja de datos o la tabla de datos y hacer cortes a lo largo y a lo ancho, para mostrar las diferentes facetas del cubo tridimensional. El hecho de no tener esta característica, complica un poco la presentación de los resultados, pues se necesitaría un especialista en sistemas que pudiera acomodar las tablas de datos, para analizar diferentes variables simultáneamente, que pueden

estar no sólo en tablas diferentes, que no se necesitan en ese momento, sino que tienen diferentes formatos y habría que procesar de nuevo la información, lo cual produce el doble manejo de información, con una resultante de mayor inversión de tiempo, esfuerzo y dinero. Para mayor comprensión del reporte de cubo, a continuación está un esquemático con la representación en 3 dimensiones, en donde cada cuadrícula representa un dato el cual puede extraerse, no importando en qué área de la base de datos se encuentre, sin tener que reorganizar las tablas y sus relaciones.



Representación del cubo de datos

b) Data Mining (Minado de datos)

Es una de las características dentro de los sistemas de información que permite al usuario final, utilizar sus datos colectados, en una forma completa, con técnicas analíticas poderosas para descubrir información que se puede transformar para predecir resultados, especialmente recomendado para la toma de decisiones. La idea que hay detrás de esta característica, es convertir los datos crudos en información para ganar terreno en mercadeo de riesgo.

Otro aspecto importante en el minado de datos es poder identificar problemas complejos en los negocios especialmente, después de aplicar las técnicas de análisis

y obtener los resultados, donde será más fácil entender el pasado e identificar los problemas que pueden venir en las proyecciones futuras.

Data Mining, **la extracción de información oculta y predecible de grandes bases de datos**, es una poderosa tecnología nueva con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información (Data Warehouse). Las herramientas de Data Mining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información (knowledge-driven). Los **análisis prospectivos** automatizados ofrecidos por un producto así van más allá de los eventos pasados provistos por herramientas retrospectivas típicas de sistemas de soporte de decisión. Las herramientas de Data Mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas y a los cuales los usuarios de esta información casi no están dispuestos a aceptar. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas.

Muchas compañías ya coleccionan y refinan cantidades masivas de datos. Las técnicas de Data Mining pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea (on-line). Una vez que las herramientas de Data Mining fueron implementadas en computadoras cliente servidor de alta performance o de procesamiento paralelo, pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como, "¿Cuáles clientes tienen más probabilidad de responder al próximo mailing promocional, y por qué? y presentar los resultados en formas de tablas, con gráficos, reportes, texto, hipertexto, etc.

Las técnicas de Data Mining se fundamentan en el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. Data Mining

toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva. Data Mining está lista para su

aplicación en la comunidad de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras:

- Recolección masiva de datos
- Potentes computadoras con multiprocesadores
- Algoritmos de Data Mining

Las bases de datos comerciales están creciendo a un ritmo sin precedentes. Un reciente estudio del META GROUP sobre los proyectos de Data Warehouse encontró que el 19% de los que contestaron están por encima del nivel de los 50 Gigabytes, mientras que el 59% espera alcanzarlo en el segundo trimestre de 1997. En algunas industrias, tales como ventas al por menor, estos números pueden ser aún mayores. MCI Telecommunications Corp. cuenta con una base de datos de 3 terabytes + 1 terabyte de índices corriendo en MVS sobre IBM SP2. La necesidad paralela de motores computacionales mejorados puede ahora alcanzarse de forma más costo - efectiva con tecnología de computadoras con multiprocesamiento paralelo.

Los componentes esenciales de la tecnología de Data Mining han estado bajo desarrollo por décadas, en áreas de investigación como estadísticas, inteligencia artificial y aprendizaje de máquinas. Hoy, la madurez de estas técnicas, junto con los motores de bases de datos relacionales de alta performance, hicieron que estas tecnologías fueran prácticas para los entornos de data warehouse actuales.

El nombre de Data Mining ha tenido un alcance que se deriva de las similitudes entre buscar valiosa información de negocios en grandes bases de datos - por ej.: encontrar información de la venta de un producto entre grandes montos de Gigabytes almacenados - y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar exactamente donde residen los valores. Dadas bases de datos de suficiente tamaño y calidad, la tecnología de Data Mining puede generar nuevas oportunidades de negocios al proveer estas capacidades:

- **Predicción automatizada de tendencias y comportamientos.** Data Mining automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos. Un típico ejemplo de problema predecible es el marketing apuntado a objetivos (targeted marketing). Data Mining usa datos en mailing promocionales anteriores para identificar posibles objetivos para maximizar los resultados de la inversión en futuros mailing. Otros problemas predecibles incluyen pronósticos de problemas financieros futuros y otras formas de incumplimiento, e identificar segmentos de población que probablemente respondan similamente a eventos dados.
- **Descubrimiento automatizado de modelos previamente desconocidos.** Las herramientas de Data Mining barren las bases de datos e identifican modelos previamente escondidos en un solo paso. Otros problemas de descubrimiento de modelos incluye detectar transacciones fraudulentas de tarjetas de créditos e identificar *datos anormales* que pueden representar errores de tipeado en la carga de datos.

Las técnicas más comúnmente usadas en Data Mining son:

- **Redes neuronales artificiales:** modelos predecible no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica.
- **Arboles de decisión:** estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen Arboles de Clasificación y Regresión (CART: Classification And Regression Tree) y Detección de Interacción Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection)
- **Algoritmos genéticos:** técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.
- **Método del vecino más cercano:** una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases.
- **Regla de inducción:** la extracción de reglas if-then de datos basados en significado estadístico.

Muchas de estas tecnologías han estado en uso por más de una década en herramientas de análisis especializadas que trabajan con volúmenes de datos relativamente pequeños. Estas capacidades están ahora evolucionando para integrarse directamente con herramientas OLAP y de Data Warehousing.

La capacidad de trabajo de Data Mining se basa en la técnica usada para realizar estas hazañas que en Data Mining se llama *Modelado*. Modelado es simplemente el acto de construir un modelo en una situación donde usted conoce la respuesta y luego la aplica en otra situación de la cual desconoce la respuesta. Por ejemplo, si busca un galeón español hundido en los mares lo primero que podría hacer es investigar otros tesoros españoles que ya fueron encontrados en el pasado. Notaría que esos barcos frecuentemente fueron encontrados fuera de las costas de Bermuda y que hay ciertas características respecto de las corrientes oceánicas y ciertas rutas que probablemente tomara el capitán del barco en esa época. Usted nota esas similitudes y arma un modelo que incluye las características comunes a todos los sitios de estos tesoros hundidos. Con estos modelos en mano sale a buscar el tesoro donde el modelo indica que en el pasado hubo más probabilidad de darse una situación similar. Con un poco de esperanza, si tiene un buen modelo, probablemente encontrará el tesoro.

Este acto de construcción de un modelo es algo que la gente ha estado haciendo desde hace mucho tiempo, seguramente desde antes del auge de las computadoras y de la tecnología de Data Mining. Lo que ocurre en las computadoras, no es muy diferente de la manera en que la gente construye modelos. Las computadoras son cargadas con mucha información acerca de una variedad de situaciones donde una respuesta es conocida y luego el software de Data Mining en la computadora debe correr a través de los datos y distinguir las características de los datos que llevarán al modelo. Una vez que el modelo se construyó, puede ser usado en situaciones similares donde usted no conoce la respuesta.

Un servidor multidimensional OLAP permite que un modelo de negocios más sofisticado pueda ser aplicado cuando se navega por el data warehouse. Las estructuras multidimensionales permiten que el usuario analice los datos de acuerdo a como quiera mirar el negocio resumido por línea de producto, u otras perspectivas

claves para su negocio. El server de Data Mining debe estar integrado con el data warehouse y el servidor OLAP para insertar el análisis de negocios directamente en esta infraestructura. A medida que el data warehouse crece con nuevas decisiones y resultados, la organización puede "minar" las mejores prácticas y aplicarlas en futuras decisiones.

5. Compatibilidad entre herramientas de Software

a) Generalidades de Excel 97

Como otro producto de la casa Microsoft, Excel 97 es otra herramienta a disposición del usuario final de computadoras, que está orientado a tener más que una información almacenada, en tenerla distribuida para hacer cálculos matemáticos, con habilidades dinámicas. Normalmente se le conoce como el manejador de datos por medio de hojas electrónicas. Esto vendría a reemplazar el concepto de "hojas tabulares", las cuales en el pasado, literalmente era la distribución de datos sobre una hoja de papel lo suficientemente ancha para poder combinar varias columnas con cálculos en una forma horizontal y vertical.

Se le ha denominado hoja electrónica, pues en lugar de una hoja de papel, ahora se utiliza un computador, con despliegues tabulares de columnas verticales, representadas por letras, que se deslizan de izquierda a derecha, en algunas versiones de hasta 256 columnas, con la habilidad de poder ampliar cada una según el ancho del dato que va en cada celda. En el sentido horizontal se encuentran las filas, representadas por números, en algunas versiones de hasta 8,182 filas, también permita ampliarse de arriba abajo, para acomodar diferentes altos de estilos de datos.

Las celdas, son los puntos de intersección o coordenadas entre columnas y filas, que es precisamente el lugar donde se pone el dato. Por ejemplo, se dice que determinado dato está en la celda A5, lo que es igual a decir que está en donde se juntan la columna A y la fila 5.

Una vez se tiene un dato que se desea calcular, se pueden emplear funciones lógicas, matemáticas, trigonométricas, estadísticas, de texto, de fecha, etc. o

bien utilizar fórmulas para combinar algún cálculo con otra celda o celdas. La parte elocuente de estos cálculos, es de que los resultados son dinámicos, pues los totales o valores ingresados para cálculo, pueden ser afectados al variar alguna otra celda de la cual se hace referencia.

Lo que hace versátil a este programa, es la amplitud de funciones que le permiten al usuario final, combinar no sólo la tarea primordial de ser una hoja de cálculo, sino combinar las funciones adicionales como las estadísticas o de bases de datos, para apoyar los resultados. Sin embargo el hecho de que se presentan estas funciones, no significa que el programa es el más adecuado para los análisis estadísticos de muchas variables o para el análisis de datos masivos, como tampoco sería el programa más adecuado para mantener una base de datos como lo hace Access 97. Sin embargo estas características son de muy buen apoyo para algunos procesos estadísticos, ya que incorpora toda una gama de formatos para gráficas, muy fáciles de usar y que no requieren mucho tiempo para organizar tareas cuya complejidad en la combinación de variables no sea mucha.

Otro punto importante de la versatilidad de Excel es la compatibilidad de sus formatos con otros programas de otras casas productoras de programas utilitarios. Con mayor razón habrá compatibilidad con los programas de la misma casa Microsoft, tal es el caso de los programas que forman parte del paquete integrado Microsoft Office, al cual pertenece Access y otros.

b) Generalidades de Access 97

De acuerdo a Microsoft, la empresa que produce este programa, Access tiene como objetivo manejar los datos del usuario a través del uso de proyectos o bases de datos, los cuales utilizarán tablas de información, con la habilidad de ordenar esos datos, por medio de una clasificación que se le conoce como registro. Este registro tendrá la información específica de un artículo en particular. Para una mejor clasificación, cada registro tiene a su vez uno o varios sub grupos de organización llamados campos. Estos campos tendrán el nivel más preciso para cada tipo de dato. Por ejemplo si el dato es de tipo numérico, todo ese campo adoptará las características que permiten que ese

dato pueda ser utilizado para cálculos matemáticos en el momento que sea necesario, no así un dato de tipo alfabético que sólo se necesita para representar una parte descriptiva de un registro, el cual no será utilizado para cálculo.

Un campo específico de una determinada tabla, también puede estar relacionado con otro campo de otra tabla dentro del mismo proyecto. Esto permite unificar información que pueda estar almacenada en diferentes lugares y que en cierto momento un reporte impreso o una consulta desee consolidar.

Access 97, permite que el banco de datos pueda tener la información necesaria para el usuario final, de una manera más ordenada, pero lo que se debe resaltar, es que el acceso a cada registro no tiene que ser secuencial, sino puede ser de acceso directo; es decir, llegar al registro que se necesita, de una forma más directa, a través de un código único de búsqueda.

Después de la búsqueda directa del código y que el puntero¹⁰ se haya ubicado en el registro que coincide con el dato que se está buscando, ya queda a disposición del usuario, la vista del campo o campos que componen dicho registro.

Access 97, viene a ser una herramienta muy útil y versátil, para aplicaciones tales como mantener las fichas o registros de usuarios, estudiantes, artículos, inventarios, ventas, facturas, etc., los cuales también se pueden compartir con otros programas externos, tales como Excel, Foxpro, Dbase, etc. Según la información provista por el departamento de mercadeo de Microsoft, la idea de utilizar este programa, es de que el usuario final, no tiene que saber de lenguajes de programación o ser un experto en sistemas de computación, para poder relacionar y acceder esa información almacenada, sino que por medio de menús y ayuda en línea, poder hacer búsquedas de los datos que quiere ver.

Algunos de los requerimientos mínimos, del equipo necesario, para que se pueda utilizar Microsoft Access 97, son los siguientes.

¹⁰/ Ver glosario, pág. 95

- Un computador personal con un procesador 486 o más reciente
- Sistema operativo o Microsoft Windows NT, o Microsoft Windows Workstation 3.5 o superior
- 16 MB de memoria RAM
- 60 MB de espacio libre en el disco duro.
- Sistema de video de señal VGA o más alto, de preferencia que maneje 256 colores
- Un ratón electrónico para poder accesar de una forma más eficiente los Odiferentes menús con opciones.

IV. ANALISIS METODOLOGICO DEL MODELO DE EVALUACION

Según el autor James A. Senn en su libro *Análisis y Diseño de Sistemas de Información*, hablando sobre el análisis de sistemas existentes, menciona: *"El objetivo del análisis de sistemas es comprender situaciones, no resolver problemas"*.

Por tanto, los buenos analistas hacen hincapié en la investigación y el cuestionamiento para conocer cómo opera el sistema e identificar los requerimientos que tienen los usuarios para modificarlo o proponer uno nuevo. Sólo después de comprender en su totalidad el sistema, los analistas están en posición de analizarlo y generar recomendaciones para el diseño de sistemas.

La forma en que se lleva a cabo la investigación de sistemas es la que determina si se reúne la información apropiada. A su vez, tener la información correcta influye en la calidad de la aplicación. En otras palabras, el buen diseño de sistemas, ya sea que se desarrollen con el método SDLS¹¹ o con los métodos de construcciones de prototipos¹² o análisis estructurado, **comienza con la documentación del sistema actual y el diagnóstico apropiado de los requerimientos de los sistemas.**

A. Investigación de requerimientos básicos

Parte de la investigación, para crear este modelo, requería que se hiciera una investigación de requerimientos en la empresa de investigación, pues se necesitaba documentar las características más sobresalientes de lo que existía en el momento que se pretendía llegar a tabular los datos del proyecto de San José Poaquil.

Esta investigación de requerimientos comprende:

1. Análisis de datos basados en hechos reales

Aquí son examinados los datos recopilados durante la investigación o estudio, incluyendo la documentación proporcionada por el usuario final.

Entre las técnicas utilizadas para encontrar estos hechos reales, se incluyeron las

¹¹/ Ver glosario, pág. 95

¹²/ Ver glosario, pág. 95

respuestas dadas a entrevistas, cuestionarios, la revisión de los registros, en el sitio mismo donde ha sucedido todo y una simulación de los procesos que se habían efectuado bajo Access 97 y Excel 97 y así tener un panorama general no solo de la parte teórica, pero también de la parte operativa.

2. Identificación de requerimientos esenciales.

Esto incluye las anotaciones de ciertas características que deben incluirse en el nuevo sistema y que se originan desde los pequeños detalles de la operación actual, hasta los criterios de desempeño, al momento de lo que se le conoce como producción. Es muy importante identificar los controles, tanto de administración como de operaciones, o seguimiento de los procesos.

3. Selección de estrategias para satisfacer los requerimientos.

En esta sección se determinan los métodos que son utilizados para alcanzar los requerimientos establecidos y seleccionados desde el principio. Aquí se le da respuesta a las cuatro preguntas importantes de los requerimientos básicos:

¿Cuál es el proceso básico de la empresa?

¿Qué datos utiliza o produce este proceso?

¿Cuáles son los límites impuestos por el tiempo y la carga de trabajo?

¿Qué controles de desempeño utiliza?

B. Técnica Activa de Análisis.

En el caso de la empresa de investigación, se decidió utilizar parte de la técnica activa, pues ya existía el hardware de operación, las herramientas de software y sobre todo, es importante aclarar que como en este proyecto lo que se buscaba era validar instrumentos que ayudaron a crear una línea base a la Institución, no se pudo observar procedimientos, ni resultados estadísticos. Sin embargo, los ejecutores del proyecto quedaron satisfechos con los que habían pedido originalmente. Después ellos solicitaron a la empresa de investigación, que les evaluaran para ese mismo proyecto, el impacto que se tuvo en las comunidades en donde se están desarrollando los proyectos.

Entonces aquí se hizo necesario investigar, aplicar los instrumentos validados, analizar los resultados, aplicar procedimientos estadísticos como la media, desviación estándar, sumatorias, porcentajes, correlaciones y regresiones lineales, con el propósito de resumir los datos recabados y obtener toda la información necesaria para facilitar la toma de decisiones a los ejecutores de estos proyectos de desarrollo que se ejecutaron en el municipio de San José Poaquil, en el departamento de Chimaltenango.

Para el diseño del modelo que aquí se propone, se llevaron a cabo todos los procedimientos de investigación para saber dónde había estado la falla interna, especialmente se evaluó si había existido doble manejo de información. Al llegar a esta fase del doble manejo de información, fue cuando se detectó el problema que se describe y para el cual se le propone una solución.

Se obtuvieron, desde el principio, las muestras que se utilizaron para el análisis del proyecto, así también la obtención de sus propósitos generales. Es oportuno indicar que la parte vital de esta investigación fue la entrevista. Aplicando la verdadera regla de la entrevista que dice "**conversar entre analista y entrevistado**" y "**no interrogar al segundo**".

En este caso, se utilizaron los dos métodos de entrevistas, la estructurada y la no estructurada, pues en algunos casos se necesitaba una respuesta cerrada para recabar mejor la información. Más adelante podrán observarse los resultados de la entrevista inicial.

A continuación se presenta una tabla que contiene las ventajas y desventajas de ambos métodos.

| COMPARACION DE LOS METODOS DE ENTREVISTA ESTRUCTURADA Y NO ESTRUCTURADA | | |
|-------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | Entrevista estructurada | Entrevista no estructurada |
| Ventajas | <p>Asegura términos uniformes en las preguntas para todos los entrevistados</p> <p>Fácil de administrar y evaluar.</p> <p>Evaluación más objetiva de preguntas y respuestas por parte de los que participan en la entrevista.</p> <p>Se necesita un entrenamiento limitado por parte del entrevistados.</p> <p>Se obtienen resultados con entrevistas cortas.</p> | <p>El entrevistador tiene mayor flexibilidad para cambiar los términos de las preguntas para que se acomoden mejor al entrevistado.</p> <p>El entrevistador puede ahondar en áreas que aparecen de manera espontánea durante la entrevista.</p> <p>La entrevista puede proporcionar información relacionada con áreas que en un principio no fueron tomadas en cuenta.</p> |
| Desventajas | <p>El costo de la preparación es alto.</p> <p>Es posible que los entrevistados no acepten un alto nivel en la estructura y planteamiento mecánico de las preguntas.</p> <p>El alto nivel de la estructura quizá no sea el más adecuado para todas las situaciones.</p> <p>El alto nivel de la estructura disminuye tanto la espontaneidad como la habilidad del entrevistador para seguir los comentarios durante la entrevista.</p> | <p>Uso ineficiente del tiempo por parte de los participantes.</p> <p>El entrevistador puede introducir sus propios sesgos en las preguntas o al notificar los resultados.</p> <p>Se puede obtener información ajena al problema.</p> <p>El análisis e interpretación de los resultados puede llevarse bastante tiempo.</p> <p>Se necesita más tiempo para reunir hechos esenciales.</p> |

C. Estudio de Campo en San José Poaquil

Básicamente se refiere a la revisión y análisis que se hizo de la empresa de investigación, la cual realizó un estudio para crear la línea base, para luego tener los elementos necesarios para medir el impacto de algunos proyectos que se realizaron en el municipio de San José Poaquil, el cual pertenece al departamento de Chimaltenango. Para comprender mejor el referido análisis se describe quién es la empresa de investigación, así como también qué fue lo que se investigó de las bases de datos, reportes y análisis estadístico ejecutado en este estudio. Debido a lo anterior es que surge la idea de crear un modelo que de forma fácil, rápida y sin necesidad de un programador, se puedan obtener resultados confiables que faciliten a gerentes, directores y coordinadores de proyectos, la toma eficiente de decisiones.

1. Empresa a la cual se analizó

“SOLUCIONES DE INVESTIGACIÓN ESTADÍSTICA”

Es una empresa dedicada a investigar, diagnosticar, analizar y evaluar datos estadísticos, por ejemplo de algunas organizaciones que ejecutan proyectos de desarrollo, cuyas administraciones pueden ser nacionales e internacionales.

Su visión se concretiza en ser una compañía con acreditación para apoyar, asesorar y evaluar el esfuerzo de transformación eficaz y eficiente en cuanto a prestación y venta de servicios de las diversas organizaciones Nacionales e Internacionales, a través de la misión de proveer soluciones y facilitar cambios relevantes a nivel del departamento administrativo y/o gerencial de las organizaciones, con el propósito de que en ellas se produzca, de manera efectiva, precisión y efectividad en la toma de decisiones.

También tiene dentro de sus objetivos, facultar a las organizaciones de herramientas tecnológicas y de calidad necesarias para su medición, evaluación, investigación y supervisión, para buscar su efectividad y competitividad.

Servicios profesionales de apoyo

- Investigación sobre gestión y desarrollo
- Elaboración de proyectos nacionales e internacionales
- Planeación Estratégica
- Evaluación y Seguimiento de la Calidad y Mejora
- Implementación de la Calidad y Mejora
- Diseño del Desarrollo de Calidad y Mejora
- Orientación a la satisfacción del usuario
- Investigación y Análisis Organizacional
- Encuestas, Entrevistas, Muestreos de Mercado y otros
- Talleres de Capacitación

2. Datos relevantes de las primeras entrevistas

Uno de los primeros aspectos que se necesitaba saber, era conocer el concepto general de la empresa de investigación y se nos dieron los siguientes conceptos. Algunos de ellos ya fueron explicados en la descripción inicial de este trabajo profesional.

“VISION: *Es ser una compañía con acreditación para apoyar, asesorar y evaluar el esfuerzo de transformación eficaz y eficiente en cuanto a prestación y venta de servicios de las diversas organizaciones Nacionales e internacionales, a través de”:*

“MISION: *Proveer soluciones y facilitar cambios relevantes a nivel del departamento administrativo y/o gerencial de las organizaciones, con el propósito de que en ellas se produzca de manera efectiva el crecimiento más cambio”.*

“Facultar a las organizaciones de herramientas tecnológicas y de calidad necesarias para su medición, evaluación, investigación y supervisión, buscando su efectividad y competitividad”.

“Definición de servicios generales

- *Gestionar e investigar los proyectos de desarrollo*
- *Elaborar de proyectos nacionales e Internacionales*
- *Planeación Estratégica*
- *Preparar la Evaluación y el Seguimiento de la Calidad y Mejora*
- *Ejecutar e implementar la Calidad y Mejora*
- *Diseñar el Desarrollo de Calidad y Mejora*
- *Orientación a la Satisfacción del Usuario*
- *Investigación y Análisis Organizacional*
- *Realizar Encuestas, Entrevistas, Muestreos de Mercado y otros*
- *Llevar a cabo Talleres de Capacitación”*

Definición de servicios específicos de apoyo

“Muchas de sus actividades están orientadas a la investigación, gestión, desarrollo, planeación estratégica, evaluación y diseño de proyectos que ejecutan organizaciones nacionales e internacionales, principalmente a las que se dedican a la intervención social; así también sus funciones están orientadas al diseño y ejecución de instrumentos para entrevistas, muestreos y análisis

organizacionales, con el propósito de orientar todas las acciones empresariales y/o institucionales, al servicio de calidad y a la satisfacción del usuario”.

“Por ejemplo si se necesita evaluar algún proyecto que tiene cierto tiempo de funcionamiento, pero jamás se ha hecho seguimiento del mismo, entonces nuestra labor es diagnosticar primero si cuentan con una línea base que genere los datos primarios como: tipo de población, número de familias, número de adultos, niños, ancianos etc., luego de obtener el diagnóstico se procede a la tarea de hacer mediciones sobre muestreo, así como la elaboración de los instrumentos necesarios, de acuerdo con la información que requiere la Institución. La gestión de un proyecto X es una tarea vital para el éxito del mismo. Dentro de la gestión, la planificación juega un importante rol, debido a que es en esta etapa donde se debe hacer las asignaciones de los recursos disponibles, para lo que es necesario estimar costos y plazos para el proyecto a realizar”.

“Hace algunos meses nos fue solicitado a través de una Organización no gubernamental, que se le evaluara el impacto de algunos de los proyectos que ellos ejecutan en el interior del país pero ellos no contaban con la información básica para dicha medición por lo que más adelante hago una breve descripción de los alcances y límites cuando no se cuenta con lo necesario”.

“Todo lo anterior surge a raíz de que en la actualidad, los avances tecnológicos están siendo cada vez más y más significativos por lo que éstas Instituciones requieren de una implementación apropiada para la utilización más eficiente de sus recursos, materiales, humanos y financieros, así como también la ejecución y control exitosa de sus proyectos”.

Qué es lo que más les importa en Soluciones de Investigación Estadística?

“Soluciones de Investigación Estadística, está dando una respuesta activa a esas necesidades reales, a través de la interlocución y la interacción entre los donantes, ejecutores y beneficiarios, fortaleciendo el marco institucional para la gestión urbana y extra - urbana dependiendo del sector en donde se ejecute el proyecto o actividad social, utilizando tecnología y herramientas que apoyan iniciativas que posibilitan el intercambio de conocimientos y experiencias, la

cooperación, el establecimiento de alianzas estratégicas, la articulación de acciones, impacto y la toma correcta de decisiones".

"También una de sus principales metas es dar un cambio radical en la orientación del marketing, en el cual lo primero debe ser el usuario, en el caso de proyectos pues definitivamente está de por medio el beneficiario, sin olvidar que aunque sean organizaciones de intervención social, no pueden estar al margen de la innovación tecnológica; por ello busca crear ventajas claras para sus usuarios. Otra de sus funciones es darles a conocer realmente quiénes son nuestros competidores, principalmente a los ejecutores o administradores de los proyectos, así como ayudarles a que todas sus intervenciones tengan visión de futuro".

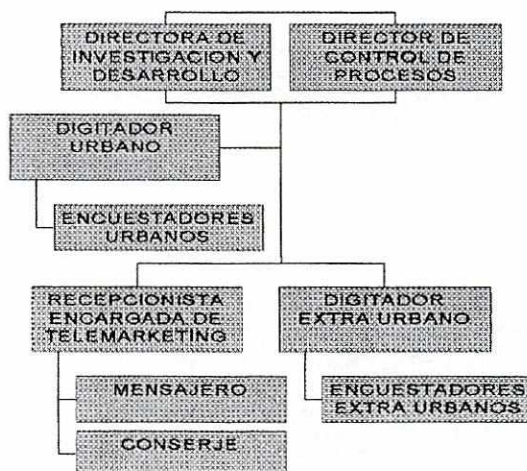
Cuál es el impacto del sistema de cómputo en la empresa?

"Definitivamente la utilización de un sistema computarizado en esta empresa, es de vital importancia y además es muy necesario, para realizar un trabajo hecho con rapidez, excelencia, máximas garantías, competitividad e impacto. Utilizar herramientas manuales significaría para nosotros incremento en los recursos humanos, financieros y materiales y sumado a ello el factor tiempo se incrementa ya que el trabajo de investigación, diagnóstico, planeación y evaluación llevaría mucho más tiempo, que si se hace uso de un eficiente sistema de cómputo, que facilite y minimice los recursos antes descritos.

Cómo está organizada la empresa

Organizacionalmente, la empresa actualmente tiene un orden jerárquico

Como se describe en el siguiente diagrama:



Cantidad de empleados al momento

“Cinco, dos directores una recepcionista, un digitador y un mensajero, los encuestadores son parte del personal subcontratado, dependiendo del volumen de trabajo”.

Las funciones de los empleados

Título del puesto de trabajo: Director de Investigación y Desarrollo

Supervisión Ejercida: Digitadores, Encuestadores, Mensajero y Conserje

Descripción del Puesto: Trabajo Profesional y especializado en el campo de la Medición, Evaluación e Investigación, coordinación y Análisis de Estudios Socioeconómicos, así como coordinación y elaboración de proyectos, validación de instrumentos, diseños muestrales, investigaciones de mercado etc.

Título del puesto de trabajo: Director de Control de Procesos

Supervisión Ejercida: Digitadores, Encuestadores

Descripción del Puesto: Trabajo Profesional en operaciones y análisis en el campo. Responsable

del área técnica de comunicaciones y sistemas de computación e integración de paquetes de software de apoyo.

Título del puesto de trabajo: Recepcionista
Supervisión Recibida: Ambos Directores
Supervisión Ejercida: Digitadores, Encuestadores, Mensajero y Conserje
Descripción del Puesto: Debe actuar dentro de un marco de responsabilidad social. Esta responsabilidad social conlleva un comportamiento ético, organización, planificación, priorización, eficacia, seguridad en la toma de decisiones y/o en las instrucciones que da. Es responsable de la organización y programación del trabajo de mensajería y conserjería, así como de la programación de trabajo de los digitadores y encuestadores. Tiene la responsabilidad de dar seguimiento a las empresas, instituciones y/u organizaciones.

Título del puesto de trabajo: Digitador
Supervisión Ejercida: Encuestadores
Descripción del Puesto: Tiene la capacitación necesaria para el ingreso de datos masivos y el conocimiento mínimo sobre algunos programas estadísticos en sistemas computarizados.

Título del puesto de trabajo: Encuestador
Supervisión Recibida: Recepcionista
Supervisión Ejercida: Ninguna
Descripción del Puesto: Trabajo técnico con alta calidad en manejo de personas y con capacidad para el manejo y comprensión de instrumentos

de medición así como el seguimiento fiel de la muestra seleccionada para la investigación.

Título del puesto de trabajo: Mensajero
 Supervisión Recibida: Recepcionista
 Supervisión Ejercida: Ninguna
 Descripción del Puesto: Trabajo técnico con la responsabilidad de llevar, recibir, organizar y controlar toda la correspondencia de la empresa, así como los pagos y cobros correspondiente y las otras que le sean asignadas.

Título del puesto de trabajo: Conserje
 Descripción del Puesto: Trabajo técnico, con la responsabilidad de mantener en condiciones optimas el ornamento, higiene y atención a las personas que están dentro como los visitantes que acuden a la Empresa.

¿Qué personal necesita utilizar una computadora?

“La directora de Investigación de proyectos supervisará aleatoriamente el ingreso de datos que hagan el digitador conjuntamente con el Director de Control de Procesos, al mismo tiempo analiza los resultados obtenidos y diseña los instrumentos y modelos necesarios de acuerdo al análisis de resultados así también tiene la responsabilidad de aplicar todas las técnicas estadísticas necesarias y los procesos de investigación, para el logro de un trabajo eficiente. El director de control de procesos estará implementando los sistemas de información computarizados e implementando programas que manejen eficientemente los datos muestrales, la recepcionista contacta usuarios vía telefónica y a través de fax, utiliza la computadora para hacer cartas, cotizaciones, informes transcripción de encuestas e instrumentos de evaluación diseñados por los directores, así como su impresión y multiplicación. En algún momento se dedica también a la digitación de datos, como una tarea adicional. Los digitadores compartirán el uso de una computadora, con la recepcionista para el ingreso de datos y graficación de los mismos, los encuestadores en algún

momento podrían ingresar datos desde la computadora portátil y luego enviarlos a la oficina central".

D. Descripción de la forma como surge el Problema

Anteriormente se hizo una breve descripción de las actividades que la empresa realiza y se mencionó sobre la línea base creada de una comunidad en Chimaltenango, a través de la validación de instrumentos de medición.

En esta comunidad se está llevando a cabo una gran serie de proyectos, a los cuales no se les había podido hacer una medición de impacto, por falta de medidas técnicas y profesionales que lo llevaran a cabo.

Es ahora cuando ya se tienen llenos los instrumentos con la información recabada; al momento no se contaba con un modelo de evaluación de datos masivos en computadora, así que utilizaron las herramientas de programas de ayuda al usuario como Excel y Access, para capturar y procesar los datos de la manera más clara posible y entregar a los coordinadores de los programas, los elementos y resultados correctos para una toma de decisiones, relacionadas con las preguntas de investigación. Los donantes, administradores y beneficiarios coincidieron en varios aspectos del proyecto evaluado, como son: su finalidad, su administración, su percepción y cómo puede mejorarse en vistas a lograr un buen nivel de impacto.

La finalidad de los proyectos que se ejecutaron en San José Poaquil están encaminados a proporcionar ayuda alimentaria dirigida a fomentar el crecimiento y desarrollo económico de los desplazados, refugiados, viudas y huérfanos de la guerrilla, pero sobre todo, crear mejores condiciones de vida para estas familias.

Respecto de lo anterior, se piensa que es de suma importancia crear un modelo que garantice óptimos resultados en un tiempo más corto y con resultados más confiables, estadísticamente comprobables y que este trabajo sea evaluado constantemente por especialistas independientes a la Institución con el propósito de que puedan proporcionarles puntos de vista críticos y desligados de intereses particulares, así ayudar a reorientar eficientemente las políticas y objetivos actuales, en busca de lograr un impacto fuerte en las mismas.

Donantes y administradores piensan que si se logra tener dicho modelo, ellos podrán incluso ampliar sus proyectos de intervención social.

Esta institución se convierte así en parte determinante de la paz y bienestar civil, al promover y proveer los insumos necesarios para alcanzar un mejor nivel de calidad de vida a través de proyectos de intervención en varios departamentos de Guatemala; siendo uno de ellos Chimaltenango.

El municipio de San José Poaquil está conformado políticamente por 1 pueblo 5 aldeas y 24 caseríos. Posee una extensión territorial de 100 kilómetros cuadrados con un clima frío. Los límites territoriales son: al Norte con Joyabaj (Quiché); al Este con San Martín Jilotepeque y San Juan Comalapa (Chimaltenango); al Sur con San Juan Comalapa, Guatemala y Santa Apolonia. Cuenta con un universo total de 15,808 habitantes, de los cuales 14,482 son indígenas, 600 no indígenas y 366 ignorados. El idioma indígena predominante es Cakchiquel, su producción agropecuaria: maíz, trigo, frijol, caña de azúcar, café, tomate, miltomate y yuca.

Tanto la Institución Ejecutora como la Donante, van en la búsqueda de tres aspectos importantes: a) alfabetización, b) reforestación y c) desarrollo humano integral.

El objeto de estudio específicamente fue el Programa de Seguridad/Ayuda Alimentaria y lo que perseguía era responder a las siguientes preguntas de investigación:

¿Cuál es la finalidad del Programa Seguridad/Ayuda Alimentaria?

¿Cómo beneficia el Programa Seguridad/Ayuda Alimentaria a las Comunidades?

¿Cómo se percibe el Programa Seguridad/Ayuda Alimentaria?

¿Cómo puede mejorar el Programa Seguridad/Ayuda Alimentaria para lograr impacto?

1. Técnicas, procedimientos y selección de la muestra:

Por las características tan peculiares del presente estudio, se llevó a cabo una investigación cualitativa al utilizar el método de triangulación. Este método crea un contraste de datos e informaciones, obtenidos por las diversas formas de investigación, así como las opiniones directas de las personas participantes y beneficiadas en el Programa Seguridad/Ayuda Alimentaria.

Además, porque es un método de recolección de información que permite comprobar y revisar su propia perspectiva a partir de datos más completos. Este es un método de crítica epistemológica que se refiere al estudio crítico del desarrollo de métodos y resultados de las ciencias. Por ejemplo en el caso del estudio que se está analizando, se logró hacer un cruce entre lo que piensan donantes, ejecutores y beneficiarios respecto de la percepción que ambos tienen acerca de los proyectos y el tipo de ayuda que se proporciona, ya que enfrenta las interpretaciones de los hechos y fenómenos extraídos de la percepción y experiencia de todas las partes involucradas en el programa de ayuda.

La información se obtuvo en las entrevistas individuales con los donantes; los responsables de la institución ejecutora del Programa de Seguridad/Ayuda Alimentaria y entrevistas grupales con los líderes comunitarios de los diversos proyectos que se están desarrollando en San José Poaquil.

También se realizaron consultas bibliográficas, documentales y cartográficas en la biblioteca de la Facultad de Agronomía de la Universidad de San Carlos de Guatemala, Unión Europea y Cáritas Arquidiocesana.

La muestra se seleccionó intencionalmente, de acuerdo a la institución y cargo que desempeñan las personas en las diferentes instituciones donantes y ejecutora. La unidad primaria de muestreo es el municipio de San José Poaquil, la unidad de análisis está centrada en los centros en donde se ejecutan los diversos proyectos de desarrollo, la unidad informativa son los beneficiarios de cada proyecto y respecto de los beneficiarios, se seleccionaron 35 líderes comunitarios y 15 participantes de cada uno de los 65 diferentes proyectos que actualmente se están ejecutando, en aldeas, caseríos y cabecera municipal en San José Poaquil. La metodología empleada para la selección de la muestra fue a través del muestreo fijo, que establece un número estable del tamaño de la muestra de tal suerte que hubo representatividad de todos los proyectos que se ejecutaron en el referido municipio, además puede decirse que la muestra fue confiable debido a que en muchos de los proyectos, se tomó en cuenta a la mayoría de los beneficiarios. La tabla que a continuación se presenta describe la muestra por género y tipo de localidad.

TABLA DE CARACTERISTICAS DE LA MUESTRA

| CARACTERISTICAS TIPO DE COMUNIDAD | Género que predomina en el liderazgo | | CANTIDAD DE BENEFICIARIOS | TIPO DE PROYECTO |
|--------------------------------------|--------------------------------------|---|---------------------------|---------------------------------------------------------------------------|
| | F | M | | |
| CABECERA MUNICIPAL | 22 | 2 | 835 | Viveros, hortalizas, reforestación, viveros de café. |
| ALDEA | 4 | 2 | 381 | Viveros, hortalizas, reforestación de ciprés y granilea, viveros de café. |
| CASERIO | 4 | 1 | 251 | Viveros, hortalizas, reforestación, viveros de café y granadía. |

2. Momento en que surge el problema

Al tener los datos iniciales se pensó en utilizar Excel como primera alternativa para almacenar la información de las encuestas, pensándose en las tabulaciones que habría que hacer y en las gráficas estadísticas que se necesitarían presentar. Luego se notó que había necesidad de relacionar algunas variables a través de todo el proyecto, pues habían códigos tales como Departamento, Municipio, Aldea, Caserío y códigos de otras variables como el tipo de proyecto, escolaridad del beneficiario, número de integrantes de la familia, tipo de vivienda, servicios públicos con los que cuentan, tipo de empleo y otras.

Es ahí cuando se decide utilizar Access 97¹³, pues tenía algunas características como mayor facilidad para crear formularios, pantallas de captura, generador de resultados, búsquedas rápidas de datos, las cuales no estaban en Excel con esa versatilidad que ofrecía Access 97. Se crearon varias pantallas para la captura de los datos que estaban en las encuestas. Al momento no había problemas, pues la captura de los datos se hacía en una forma muy simple y ordenada. Es a partir de este momento que la Empresa de investigación se enfrenta a un grado

¹³/ Ver anexo Estudio de Campo en Access 97. Pág. 98-102

de dificultad mayor, por la forma en que se diseñaron las tablas iniciales en Access 97. Esta evaluación requería varios cruces de variable que se encontraban en diferentes tablas y diferentes proyectos. El grado de dificultad aumentó cuando las relaciones entre las tablas eran totalmente diferentes a las que existían al momento. En ciertas ocasiones solamente se necesitaba un dato de alguna tabla para poder hacer otras combinaciones con otras variables de otras tablas, analizar el comportamiento de dicha variable, someter esos datos a análisis estadísticos computarizados, para producir un resultado y posteriormente ser interpretados por el analista estadístico para tomar una decisión. Para llegar a este paso había que reconstruir una tabla existente, con los campos de otra tabla y así sucesivamente. Si el analista estadístico necesitaba hacer otras combinaciones, el trabajo de crear y reconstruir tablas se iniciaba de nuevo. Ahora bien, para poder hacer las presentaciones de resultados se necesitaba un procedimiento similar. Indudablemente había un exagerado doble manejo de información, que redundaba en tener que emplear a un experto en sistemas de computación, especialmente en el área de Access, combinar diferentes programas, ej. Uno para combinar campos, otro para hacer análisis estadístico y otro para presentar la información escrita y con gráficas, no sin mencionar el demasiado tiempo para llegar a hacer el análisis por variables, costos elevados y lo que era más peligroso, poder alterar un dato que fuera a sesgar los resultados estadísticos.

De acuerdo al análisis de los aspectos relevantes en la entrevista y la revisión de información recabada con los instrumentos se piensa que el problema tuvo su origen en la selección del software que se tenía pensado para tomar los datos de los instrumentos y luego la aplicación de procedimientos estadísticos.

Primero se utilizó una hoja electrónica, pues allí se habían hecho otras tabulaciones y análisis para otros proyectos y funcionó apropiadamente, pero no eran tantos datos ni con diversidad de variables y grupos.

En una de las entrevista, el director de estadística mencionó que era más necesario, tener un sistema de análisis estadístico, que un programa que solamente manejara una base de datos.

Este programa debiera hacer análisis de más de tres variables y que además que se puedan obtener análisis cruzados y combinaciones múltiples en el menor tiempo, para hacer simulaciones. Así también, si es posible, no necesariamente tener que depender de un experto en computadoras o programador, pues complica aún más la situación en el aspecto económico y lo que es más, la diferencia de terminología y conocimientos en ambas partes; la de sistemas de computación y la de estadística.

E. Descripción del hardware utilizado en la empresa investigada

Las características del equipo utilizado son las siguientes:

1. Computadoras

1 Computadora sin marca, de tipo compatible, sin conexión a red con los siguientes componentes:

- 64 Mb. de memoria Ram,
- Procesador Pentium II de 500 de 266 Mhz
- Disco Duro con capacidad de 10 Gb.
- Monitor SVGA a colores de 14"
- Fax módem de 56 Kbps .
- Sonido y bocinas.
- Dispositivo para diskettes de 1.44 Mb.
- Teclado para windows y mouse compatible
- Puertos de comunicaciones paralelo y seriales.
- CD ROM 40X

1 Computadora portátil sin conexión a red, con los siguientes componentes:

- 16 Mb. de memoria Ram,
- Procesador Intel Pentium de 120 Mhz
- Disco Duro con capacidad de 1.0 Gb.
- Pantalla a color de tipo pasivo de 12"
- Fax módem de 56 Kbps .
- Sonido y bocinas.
- Dispositivo para diskettes de 1.44 Mb.
- Teclado incorporado y mouse compatible
- Puertos de comunicaciones paralelo y seriales.
- CD ROM 16X

2. Impresoras

1 Impresora marca Epson Stylus 600, con los siguientes componentes:

- Impresión a colores y negro
- Método de impresión por inyección de tinta
- Para tamaños de papel carta e individuales

F. Software utilizado por la empresa de investigación

A continuación se mencionan los programas, relacionados con el presente modelo, que la empresa de investigación utilizó cuando como apoyo para tabular los datos y obtener los resultados estadísticos del proyecto de San José Poaquil.

1. Sistema Operativo

DOS con Windows 98

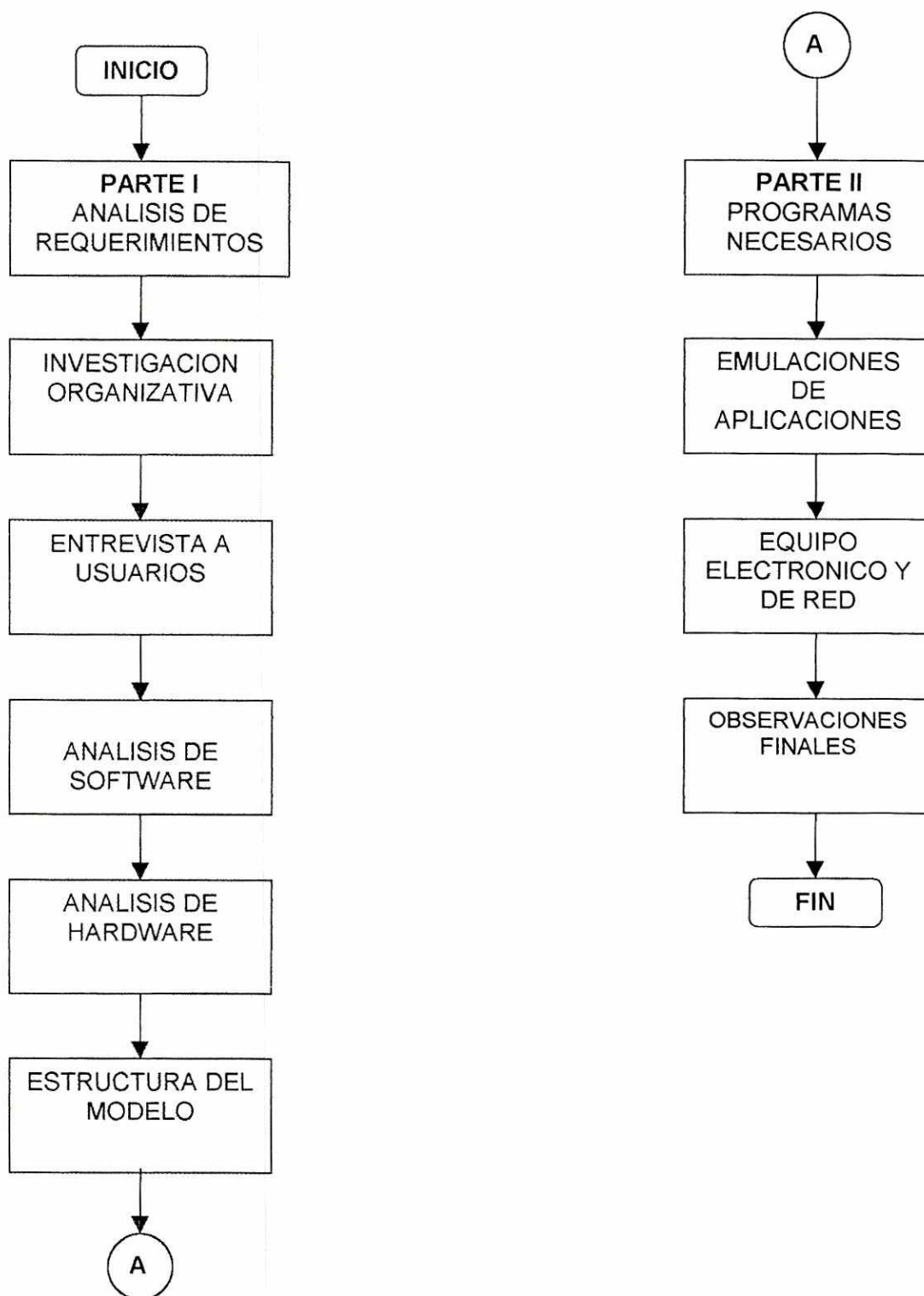
2. Programas de apoyo al usuario

Excel 97

Access 97

A continuación se presenta el esquema del modelo computarizado de análisis y evaluación de datos, para una mejor comprensión de todo el análisis metodológico empleado.

ESQUEMA DEL MODELO COMPUTARIZADO DE ANALISIS Y EVALUACION DE DATOS,
CON RESULTADOS ESTADISTICOS



V. ESTRUCTURA DEL MODELO

Después de que se aplicaron los pasos importantes en el análisis de sistemas, tales como determinar los requerimientos básicos, los requerimientos de las transacciones de los usuarios, requerimientos de toda la organización, las muestras de salidas y de entradas, etc., se aplicó un paso más y fue la participación del usuario en el momento del diseño del modelo. En relación a este paso, menciona :James A. Senn:365: en su libro *Análisis y Diseño de Sistemas de Información* describe, "Incorporar las ideas de los usuarios durante el proceso del diseño y desarrollo, nos permitirá ver que tan bien trabaja el diseño en una situación de la vida real". Y efectivamente, los usuarios finales, que son los del problema a resolver y que llaman al analista, tienen la información que el mismo analista y/o diseñador necesitan. Los usuarios son los que conocen sus áreas específicas de trabajo mucho mejor que el que llega a investigar únicamente. Qué mejor que combinar la experiencia e información del usuario con la del analista para diseñar algo que se está probando en la vida real y que satisfaga las necesidades del usuario final.

Después de esta identificación de los requerimientos de los sistemas de información y las especificaciones de alternativas de diseño, también se agregaron paralelamente los criterios más relevantes observados a lo largo del análisis. Estos criterios importantes fueron los siguientes:

- Utilizar hasta donde fuera posible, los recursos de hardware, que recién habían sido adquiridos por la empresa de investigación. Si para cumplir el objetivo se necesitaba adquirir más equipo, tratar de regirse a un presupuesto para estudio.
- Utilizar las herramientas de software con las que contaban al momento, sin descartar la posibilidad de la adquisición de nuevos programas, ya que se consideraba que en esta área se habían dado la mayor parte de casos que contribuyeron al doble manejo de información. También regirse a un presupuesto para estudio.
- Que el diseño de la solución, a través de este modelo, quede dentro de un esquema de simplicidad operatoria para los usuarios finales.
- Que el software en general y el software de aplicación estén enfocados más en un sistema de análisis estadístico y no solamente en un sistema de cómputo elaborado que tenga capacidad para manejar datos masivos. Esto no significa que el grupo de programas no pueda llenar las otras necesidades como tareas de tipo secretarial y de administración.

- Que al diseñar el sistema completo, tanto en el área de hardware como de software, quede como una herramienta para los estadísticos, flexible y práctica sin que tengan que depender de un ingeniero en sistemas de tiempo completo o de un operador especializado en sistemas de cómputo. La función de un ingeniero en sistemas puede quedar flexible como una consultoría, cuando así se necesite.
- La solución tanto física como lógica, debe contemplar el tamaño de la empresa actual, más un margen de crecimiento de la misma.
- En resumen debe ser un sistema aplicado, regido por las reglas de mantenimiento simple.

A. Programas necesarios en el modelo

Esta sección comprende tres divisiones, las cuales son las que forman toda la plataforma lógica sobre la cual están descansando las cuatro funciones que proveerán los resultados de cada proyecto; estas funciones son: entradas, procesos, salidas y almacenaje.

Las divisiones son: Sistema Operativo y Enlace de Red, Programas de Aplicación y Programas de soporte.

1. Sistema Operativo y Enlace de Red

Cuando se obtuvieron los requerimientos y se efectuó el análisis, un punto determinante para sugerir el sistema operativo y enlace de red, fue el que a corto y mediano plazo solamente un usuario estará accediendo los datos burdos y también los organizados, para ser analizados estadísticamente en un computador. Este usuario es el único en la empresa de investigación que decide cuáles son las variables a utilizar para analizar y hacer simulaciones y luego tomar decisiones de los resultados a presentar y también cómo presentarlos. Sin embargo, este usuario también necesitará acceder la información que originalmente se ingresó en otra computadora de otra oficina. Esta información pudo haber sido capturada por un digitador en otro computador, por la secretaria en su computador personal o por el coordinador de procesos en su estación de trabajo o bien pudo llegar desde un punto remoto por medio de una transmisión vía módem, en uno de los formatos del programa de soporte que se esté utilizando.

Dado a que el flujo de datos va en su mayor parte en una dirección a un punto central para análisis y lo que se busca es más que todo compartir algunos recursos físicos y

lógicos, no es necesario pensar en tener un computador que esté como servidor central dedicado, con un sistema operativo como Windows NT, Windows Workstation, Novell Network, Linux, etc., lo que involucraría invertir en un recurso que puede ser muy útil en el futuro, pero que no es la prioridad al momento, según la estructura y el presupuesto actual de la empresa.

Sin embargo deben quedar provista una estructura por si se diera el caso de hacer una conversión futura donde estará un servidor dedicado. Esta conversión, cuando exista, debiera ser sin que involucre tener que hacer un nuevo diseño completo tanto físico como lógico.

Los dos computadores que están actualmente funcionando, ya poseen sistema operativo con sus respectivas licencias. La que es utilizada en análisis estadístico cuenta con Sistema Operativo Microsoft Windows 98, segunda edición, la portátil, ya cuenta también con Microsoft Windows 95. Para las otras dos que están contempladas en el diseño del modelo, las cuales son explicadas más adelante, se sugiere que se utilice Microsoft Windows Millennium, el cual es el más actualizado para este tipo de enlace y proyecto, ya que al momento de escribir este modelo, Windows 98 está saliendo del mercado.

A pesar de que son diferentes sistemas operativos entre los equipos actuales y los que se están sugiriendo adicionalmente, pertenecen a la misma familia o plataforma y al hacer pruebas de conexión, en las instalaciones de la empresa de investigación, todos los equipos funcionaron apropiadamente.

Lo ideal es que todas las computadoras posean el mismo sistema operativo, pero en esta situación para estandarizar dicho sistema, se requiere también de cambios en algunos componentes de electrónica y configuración. Por ejemplo antes de pensar en hacer una actualización de licencia de Windows 95 a Windows Millennium, en la computadora portátil, sería ideal aumentar de 16 Mb. de memoria central a 32 Mb. o 64 Mb.

En la computadora fija, la cual ya existe en la oficina de análisis, solamente sería de actualizar la licencia de Windows 98 a Windows Millennium. Todas estas actualizaciones, por supuesto tendrían un impacto en el aumento del presupuesto

previsto para este modelo. Cuando se propuso este cambio, la empresa de investigación sugirió hacerlo más adelante.

2. Programas de Aplicación

Se le denominó "*Programas de aplicación*", para hacer una diferencia entre los programas que componen el sistema operativo, los programas de apoyo al usuario, éste y/o estos programas que están dedicados exclusivamente a brindar las herramientas y llenar las necesidades de mayor prioridad de la empresa; en este caso la Estadística. Tal y como se ve en el marco conceptual de este modelo, la mayor jerarquía, de esta empresa, es la de realizar tareas para servicio al usuario, enfocadas en el análisis e interpretación de resultados de datos estadísticos.

En este aspecto es en donde la empresa de investigación manifestó su mayor limitación y mala utilización de los recursos (hardware, software, humanos, tiempo, costos y el doble manejo de información); tal y como se manifestó en la sección que describe el problema primordial de la empresa de investigación.

Se invitó al representante de SPSS de Guatemala y Centro América, para hacer una presentación del programa base y tablas, con el propósito de analizar detalladamente si dicho programa respondía a las necesidades y jerarquías de la empresa de investigación, así como también si respondía a los requerimientos establecidos por las organizaciones locales e internacionales, que son los que dirigen y ejecutan los proyectos de desarrollo e investigación. La presentación se llevó a cabo en las instalaciones de la empresa de investigación, estando presentes el equipo profesional y técnico, al igual que la persona que efectuó la investigación para la preparación del presente modelo. En dicha reunión se planteó un esquema que contenía los formularios específicos que se habían utilizado en el proyecto donde se necesitó hacer los cruces de variables y las mediciones de impacto logrados en la ejecución de los proyectos.

La base de datos que utilizó la empresa de investigación, cuando realizaron el proyecto de San José Poaquil, fue desarrollada en Access con el nombre de *Instrumentos 2*; la cual contiene las siguientes tablas:¹⁴

¹⁴/ Ver anexo : Proyectos, pág. 101

- a) *Departamentos* (contiene los campos del departamento de Chimaltenango)
- b) *Municipios* (contiene los campos relacionados a información de los municipios de Chimaltenango)
- c) *Aldeas* (contiene los campos relacionados a la información pertinente a las aldeas que componen el municipio específico de San José Poaquil, que fue donde realizó el estudio la empresa de investigación.
- d) *Caseríos* (contiene los campos relacionados con la información pertinente a los caseríos del municipio de San José Poaquil)
- e) *Tipo de Proyecto* (información de encabezado para relacionarlo con la tabla *Proyectos*)
- f) *Proyectos* (contiene la información de los diferentes proyectos que la empresa de investigación midió a través de las encuestas)
- g) *Encuesta 1* (contiene los campos de la línea basal, que soportaron a la primer encuesta en la localidad de San José Poaquil)
- h) *Encuesta 2* (contiene los campos de la segunda encuesta, donde aparece la información al detalle de las variables estadísticas a medir)

El siguiente paso fue observar las relaciones de llaves primarias entre las tablas anteriores, desde SPSS.¹⁵

Seguidamente se ejecutó el programa SPSS, que se pretende utilizar como una solución de software. Como se quiere utilizar el menor tiempo posible para probar la eficiencia de este programa, se empleó una característica para importar archivos de diferentes formatos, ej: de Oracle, Excel, Access, Informix, Foxbase, Sybase, etc. En este caso se importó la base de datos creada originalmente en

Access. Al importarla, se observó que las relaciones de sus llaves primarias fueron consistentes en la nueva estructura. De no haber sido de esa manera hay una opción para poder crear nuevas llaves o nuevas relaciones.

¹⁵/ Ver anexo : Reporte de cubo con tabla pivotada, pág. 105

Ya con el conocimiento anterior de lo que se pretendía estudiar, las variables que se querían medir y los procesos estadísticos que se querían aplicar, se procedió a seleccionar las tablas, los campos, las estadísticas y las gráficas.¹⁶

Seguidamente se le dieron las instrucciones al programa de SPSS, para generar el primer reporte con los datos que el director estadístico solicitó. Estas instrucciones tomaron aproximadamente 2 minutos. Después de dar la orden de ejecutar se procesó el primer reporte de cubo con una tabla pivotada, con los datos del proyecto de infraestructura. El proceso tomó, 6 segundos¹⁷.

El tiempo y esfuerzo que se tomó para generar el primer reporte fue muy poco comparado con el que tomó originalmente cuando se utilizó Access y Excel.

Seguidamente se hicieron cambios en el área de OLAP y reportes de cubo pivotes, los cuales involucraba el arrastre de un ícono que estaba definido con una variable, hacia otro lugar del cubo de datos y se ejecutó para ver el reporte anterior bajo otros ejes¹⁸. El proceso de generar este nuevo reporte tomó aproximadamente 6 segundos.

Los siguientes reportes fueron generados a continuación:

- Proyecto de Reforestación. Reporte de cubo con una tabla estadística¹⁹
- Proyecto de Reforestación. Reporte de cubo con una tabla de frecuencias con gráficas de tipo histograma, incluyendo una curva normal²⁰
- Proyecto de Introducción de Agua. Reporte estadístico con una tabla de frecuencias, con gráficas de tipo histograma, incluyendo una curva normal.²¹

¹⁶/ Ver anexo 2: Tablas pivotadas, Pág. 108

¹⁷/ Ver anexo 2: Tabla estadística, Pág. 109

¹⁸/ Ver anexo 2: Reporte de cubo con tabla estadística, Pág. 110

¹⁹/ Ver anexo 2: Tabla de frecuencias - reforestación, Pág. 111

²⁰/ Ver anexo 2: Tabla de frecuencias - introducción de agua, Pág. 120

²¹/ Ver anexo 2: Correlaciones bivariadas - introducción de agua, Pág. 121

- Proyecto de Introducción de Agua. Reporte estadístico de frecuencias con prueba estadística no-paramétrica de Chi cuadrada²².

Los procedimientos para cambiar de un tipo de reporte a otro, tomó de 1 a 2 minutos y después de dar la orden de ejecución, tanto el procesamiento de datos y la generación del reporte se efectuó en 4 segundos. El tiempo de impresión en papel aproximadamente, 1 minuto por hoja.

En resumen: Con SPSS se logró, en aproximadamente 1 hora, al utilizar una persona, lo que con Access 97 y Excel 97 tomó aproximadamente 2 días de trabajo (20 horas) y si se quisieran nuevos reportes, tomaría mayor tiempo que si se hicieran en SPSS, además, utilizando dos personas, un analista estadístico y un especialista en sistemas de computación. Fue muy evidente y convincente, que el uso apropiado de la herramienta correcta, en este caso un programa aplicado a la Estadística, en las manos del analista estadístico evitó: el doble manejo de información, eliminar el riesgo de sesgo de datos, tener que estar en un ambiente de computación complejo y aumentar el presupuesto para tener que contratar de planta, a un especialista en computación.

Vale la pena mencionar que aparte de las observaciones anteriores, se dejaron ver todas las otras características adicionales del programa, que se empezaron a aplicar a esos datos del experimento, especialmente en lo que respecta al análisis y resultados estadísticos mucho más completos y complejos.

Estas pruebas y simulaciones que se hicieron posteriormente, fueron hechas en un tiempo impresionantemente corto y con resultados que daban evidencia de que se tenía en ese momento la herramienta correcta para este tipo de aplicación y necesidades.

Tanto los reportes como las gráficas, se muestran en su forma original, sin las modificaciones de color, títulos, fondos, tipos de letra, etc., para lo cual hay diferentes tipos de menú y ventanas para hacerlos.

²²/ Ver anexo 2: Chi cuadrada - introducción de agua, Pág. 122

- La presentación del programa fue tan convincente de sus beneficios en velocidad y flexibilidad, que se solicitó, por parte de la administración de la empresa de investigación, que el representante de SPSS, dejara una versión de tipo demo del programa, para una evaluación posterior, especialmente para analizar el contraste entre diversas variables, gráficas y reportes. A lo cual se accedió y quedó instalado y funcionando por 30 días. Durante ese tiempo hubo oportunidad de ver las ventajas de utilizar la mayor parte de funciones del programa, especialmente las que eran más importantes: el Proyecto de Introducción de Agua.
- También se capacitó al director de análisis y al director de control de procesos, por 18 horas, en el sistema base, para una mejor operación.
- Pasados los 30 días de evaluación del programa base, la administración decidió comprar ese módulo, más el módulo tablas en el modo monousuario. El programa se instaló en la computadora del director de análisis únicamente, por dos razones:
 - a) El costo para tener el sistema de SPSS en modo multiusuario (con dos licencias) elevaba mucho el costo de la inversión.
 - b) No era necesario correr el programa SPSS en más de una computadora a la vez, pues los datos se pueden capturar utilizando otros programas utilitarios, compatibles con el formato de SPSS. Tal es el caso de Excel o Access pero en su mayor parte para la captura de información y no para análisis estadístico.

3. Los programas de soporte

Estos no son más que los programas que la empresa de investigación ya poseían con licencias, como son Excel y Access. Estos dos programas fueron empleados para varias capturas de datos por el digitador y en ciertas ocasiones por la recepcionista y el director de control de procesos. Una de las ventajas que ofreció SPSS es la compatibilidad o habilidad para poder leer los formatos de estos dos programas y de otros más de casas productoras de programas varios.

A los nuevos computadores se les instaló Excel y Access, así que si en determinado momento llegara a ser necesaria la toma de datos para análisis cada estación disponía de ellos y posteriormente por medio de la red local, se podrían compartir los archivos generados por las mismas estaciones.

El otro programa que se utilizó para hacer pruebas de transmisión de datos del computador portátil al computador de la Dirección de Control de Procesos, es el conocido en Inglés como "PC Anywere", por razones de presupuesto y prioridades de la administración no se instaló en ningún computador, sin embargo se utilizarán los medios de comunicación provistos por Internet y correo electrónico al momento, para el envío y recibo de datos remotos.

Sumados a éstos, hubo otras instalaciones de programas utilitarios como detectores de virus, procesadores de palabras, programas para hacer presentaciones por medio de retro-proyector digital. A continuación se algunas de las bondades del programa estadístico SPSS.

4. Generalidades de SPSS 10.0

SPSS Inc. Por sus letras en Inglés: Statistical Package for the Social Sciences (SPSS), (Paquete Estadístico para las Ciencias Sociales), es una casa productora de software en E.U., que se ha dedicado a diseñar programas que proveen soluciones a profesionales de distintas áreas, donde se necesite hacer predicciones, para la toma de decisiones.

Estos programas se integran, especialmente para el análisis de mercado, análisis gerencial y operacional donde haya grupos de datos pequeños o grupos masivos.

Su primer producto fue desarrollado en 1968, por 3 estudiantes de la Universidad de Stanford, quienes forman actualmente parte de la mesa directiva. Este producto fue un sistema integrado con varios programas de Estadística. En 1975 los tres estudiantes terminaron SPSS Inc.

Al inicio los programas operaban fundamentalmente en sistemas de computadoras grandes, conocidas como Mainframes (Sistemas Principales) y no fue hasta en 1984 cuando introdujeron el primer sistema de estadística para computadoras personales. En

1992, SPSS llegó a ser el mayor desarrollador de software para estadística, que ofreciera un producto para Microsoft Windows y seguidamente para Windows 95 y las subsiguientes versiones. Actualmente ofrece su producto estadístico en 9 idiomas.

Entre 1994 y 1999 SPSS adquirió 9 corporaciones más que tenían vínculos en los temas principales de Investigación. Esas empresas eran: SYSTAT Inc., BMDP Statistical Software, Jandel Scientific Software, Clear Software, (Software Claro), Quantime Ltd., In2itive Technologies A/S, Integral Solutions Ltd. y Vento Software Inc.

En el año 2000, llegó a ser una de las empresas líderes en soluciones analíticas, de comportamiento y de predicción, incluyendo telecomunicaciones, cuidado de la salud, banca, finanzas, venta al detalle y venta por mayor, pero especialmente en la investigación de mercados que necesitan muestreos.

Dos características principales que tienen sus programas analíticos, son Data Mining (Minado de Datos) y OLAP (On Line Analytical Processing, Procesamiento Analítico en Línea). Características que claman ellos, son las que hacen la diferencia de otros paquetes estadísticos.

Además de las dos características anteriores, hay otras ventajas en el diseño de sus programas, pues cuentan con diferentes módulos para expandir las soluciones de diferentes necesidades. Estos agregan mayor fuerza analítica al acceso de datos, a la preparación de resultados, segmentación y modelos predictivos.

Después de entrevistar al representante de SPSS en Guatemala, Centro América y el Caribe, nos brindó la siguiente clasificación de los diferentes productos que componen SPSS, en una escala de menor a mayor venta, pensando en los usuarios con las características similares y necesidades generales. A continuación se describen en ese orden:

- SPSS Iniciador de Carreras
Especialmente para el estudio analítico en el salón de clases. Preparación académica.

- SPSS Interfaces de Acceso de Datos

Para construir una solución analítica, cuando se crean interfaces directas con productos claves en el mercado.

- **Paquete para graduandos**
Software avanzado para preparación académica.
- **SPSS Tendencias**
Mejora las predicciones, al analizar la información histórica, asimismo construye modelos y predice tendencias al usar análisis por series de tiempos.
- **TELEform Estandar**
Automatiza el proceso para convertir las encuestas en papel, en información en línea.
- **SPSS Análisis de Variables Perdidas**
Mejora las conclusiones cuando llena los espacios en blanco con sus datos extraviados.
- **SPSS Mapas**
Para descubrir cómo las variables geográficas afectan la toma de decisiones, al transformar sus datos en información demográfica.
- **SPSS Pruebas Exactas**
Para lograr conclusiones correctas, con muestras pequeñas y ocurrencias raras en bases de datos muy grandes.
- **SPSS Conjunto**
Para mejor entendimiento de las preferencias de los usuarios, especialmente para lanzamiento de nuevos productos al mercado.
- **SPSS Categorías**
Predice eventos por venir, categóricamente por ítems y mapas agrupados porcentuales
- **SPSS Modelos Avanzados**

Analiza relaciones complejas al utilizar herramientas de modelaje de alto nivel.

- **SPSS Tendencias**
Mejora las predicciones, al analizar la información histórica, asimismo construye modelos y predice tendencias al usar análisis por series de tiempos.
- **SPSS Ingreso de Datos**
Ayuda a mejorar la satisfacción del usuario, al identificar nuevas oportunidades, mide las actitudes de los usuarios, colecta la información que es necesaria.
- **SPSS Modelos de Regresión**
Procedimientos poderosos de regresión.
- **SPSS Tablas**
Generador de reportes que muestran claramente, aun cuando hay que hacer análisis de los más complicados y luego comunicar sus resultados.
- **SPSS Base**
Para la toma de decisiones, análisis estadístico, preparación de datos, reportes analíticos y modelaje de predicciones. La nueva versión 10.0, para Windows 98, ofrece la flexibilidad de acceder y analizar conjuntos de datos masivos. Exigen ellos que prácticamente elimina las limitaciones de tamaño de archivos. Contiene el motor de más de 150 procedimientos estadísticos y la generación de Gráficos en 2 y 3 Dimensiones (Interactivos), así como los cubos OLAP.

Otra de las ventajas es de que un usuario puede unir varias tablas con información y prepararlas para ser analizadas por SPSS, sin necesidad de conocer las herramientas de consulta como SQL. También maximiza los valores existentes en los archivos de Excel, al importar los datos hacia SPSS y aplicarles las Estadísticas necesarias para obtener resultados.

Dentro del manejo de OLAP y DATA MINING (minado de datos), utiliza un tipo de tecnología conocida como Pivot, que es el giro de datos sobre un eje. El usuario puede crear las tablas, cambiar filas, columnas y diferentes capas de datos para preparar las gráficas y producir reportes de cubo, no importando realmente donde están colocados los datos.

B. Equipo utilizado en el modelo

El utilizar los criterios mencionados al inicio de este capítulo permitió llegar a un punto clave para seleccionar el equipo. Sin embargo se agregaron otros factores como ver las especificaciones mínimas requeridas por los programas de aplicación, los de enlace y los de apoyo. Veamos a continuación como quedó funcionando el modelo.

1. Oficina de Dirección y Análisis Estadístico

En esta oficina es donde se llevarán a cabo los análisis y estudio de resultados estadísticos, además de las funciones de administración general. Es aquí donde se instaló el programa de aplicación SPSS Base y Tablas. Por ser el departamento que demanda mayor atención a dichos procesos, se instaló un computador más veloz y con mayores capacidades de almacenaje temporal y permanente, que el que se encontraba originalmente, el cual ya fue descrito.

1 Computador de tipo PC con las siguientes características:

- 128 Mb. de memoria Ram,
- Procesador Intel Pentium III de 750 Mhz
- Disco Duro con capacidad de 20 Gb.
- Monitor SVGA a colores de 14"
- Fax módem de 56 Kbps .
- Dispositivo para diskettes de 1.44 Mb.
- Teclado para Windows y mouse
- Puertos de comunicaciones paralelo y seriales.
- CD ROM 40X, o superior, con sonido y bocinas.
- Tarjeta de comunicaciones para red.

- Sistema interno para hacer copias de archivos, de tecnología Zip, a manera de respaldo ante una catástrofe.

Algunos características importantes que son indispensables:

- 128 Mb.de memoria, porque el requerimiento mínimo de SPSS es de 64 Mb. Tener más memoria que la requerida, es aprovechada por el procesador, para expeditar los procesos internos de entradas y salidas.
- Se instaló un sistema interno de respaldo tipo ZIP, para tener copias archivadas en cartuchos, de los datos que se están procesando de cada proyecto, en el evento de que surgiera cualquier imprevisto y se perdieran todos los datos, se podría recurrir a esta alternativa para poder recuperar los mismos.

1 Impresora, con las siguientes características:

- Impresión a colores y negro, con velocidad de 8 páginas por minuto.
- Método de impresión por inyección de tinta
- Para tamaños de papel carta e individuales

1 Sistema ininterrumpido de energía de 750 Watts.

- Capacidad para mantener el fluido eléctrico ante una falla de energía de la fuente central.
- De 5 a 20 minutos de respaldo, por medio del uso de baterías incorporadas.
- Regulador de voltaje incorporado para protección del equipo.

2. Oficina de Dirección de Control de Procesos

En esta oficina se llevan a cabo investigaciones y análisis menos complejas, también se preparan los formatos para muestreo, se coordinan las entradas de datos tanto local como remota, se verifica la integridad de los datos crudos, se preparan las pantallas de captura de datos en Excel o Access y en algunos momentos se comparte para la captura de datos.

Se instaló una impresora de alta velocidad con carro ancho, pues a veces se necesita generar listados de todos los datos ingresados para verificación o presentación al usuario. Hacerlo en una impresora laser o de inyección de tinta, sería muy alto el costo por el volumen de datos.

Aquí también se instaló un computador de similar capacidad a la que quedó funcionando en la Dirección de Análisis.

1 Computador de tipo PC con las siguientes características:

- 128 Mb. de memoria Ram,
- Procesador Intel Pentium III de 750 Mhz
- Disco Duro con capacidad de 20 Gb.
- Monitor SVGA a colores de 14"
- Fax módem de 56 Kbps .
- Dispositivo para diskettes de 1.44 Mb.
- Teclado para Windows y mouse
- Puertos de comunicaciones paralelo y seriales.
- CD ROM 40X, o superior, con sonido y bocinas.
- Tarjeta de comunicaciones para red.
- Sistema interno para hacer copias de archivos, de tecnología Zip, a manera de respaldo ante una catástrofe.

1 Impresora, con las siguientes características:

- Impresión en negro de alta velocidad, de 435 caracteres por segundo
- Método de impresión matricial
- Para tamaños de papel hasta de 14" de ancho

1 Sistema ininterrumpido de energía de 750 Watts.

- Capacidad para mantener el fluido eléctrico ante una falla de energía de la fuente central.
- De 5 a 20 minutos de respaldo, por medio del uso de baterías incorporadas.
- Regulador de voltaje incorporado para protección del equipo.

Además del equipo anterior, esta oficina, también tendrá a su cargo el computador portátil, el cual se utilizará en las presentaciones a los usuarios o para la captura de datos en áreas remotas. Se sugirió ampliar la memoria, de 16 Mb. a 64 Mb. Asimismo incorporar una tarjeta electrónica para enlace de red, para que pueda utilizarse no sólo en actividades remotas o locales, sino dentro de una conexión física a la red.

1 Computador portátil, con las siguientes características:

- 16 Mb. de memoria Ram,
- Procesador Intel Pentium de 120 Mhz
- Disco Duro con capacidad de 1.0 Gb.
- Pantalla a color de tipo pasivo de 12"
- Fax módem de 56 Kbps .
- Sonido y bocinas.
- Dispositivo para el uso de diskettes de 1.44 Mb.
- Teclado incorporado y mouse compatible
- Puertos de comunicaciones paralelo y seriales.
- CD ROM 16X

3. Oficina de Recepción y Entrada de Datos

A esta localidad se trasladó el computador que originalmente se utilizó para hacer los análisis del proyecto de investigación de este modelo.

Aquí se compartirán las tareas secretariales con la de entrada de datos. Por razones de costos e inversión no se hicieron modificaciones a la velocidad del procesador o a la memoria RAM, ya que en las pruebas que se hicieron la velocidad fue aceptable. Se incorporará un sistema interno para copia de archivos, tipo ZIP.

1 Computador de tipo PC con las siguientes características:

- 64 Mb. de memoria Ram,

- Procesador Pentium II de 500 de 266 Mhz
- Disco Duro con capacidad de 10 Gb.
- Monitor SVGA a colores de 14"
- Fax módem de 56 Kbps .
- Dispositivo para diskettes de 1.44 Mb.
- Teclado para Windows y mouse compatible
- Puertos de comunicaciones paralelo y seriales.
- CD ROM 40X, con Sonido y bocinas.
- Sistema interno para hacer copias de archivos, de tecnología Zip, a manera de respaldo ante una catástrofe.

1 Impresora marca, con las siguientes características:

- Impresión a colores y negro, con velocidad de 6 páginas por minuto.
- Método de impresión por inyección de tinta
- Para tamaños de papel carta e individuales

1 Sistema ininterrumpido de energía de 750 Watts.

- Capacidad para mantener el fluido eléctrico ante una falla de energía de la fuente central.
- De 5 a 20 minutos de respaldo, por medio del uso de baterías incorporadas.
- Regulador de voltaje incorporado para protección del equipo.
- Para tamaños de papel carta e individuales

C. Equipo para Enlace de Red

Tal y como se explicó en la sección "Sistema Operativo y Enlace de Red", se escogió la topología de red de punto a punto, en la cual todas las estaciones de trabajo pueden compartir sus recursos físicos y lógicos, sin que haya forzosamente un computador que actúe como servidor central. Cada computador en este tipo de topología viene a ser, por así decirlo, un servidor independiente que comparte sus recursos. Por lo tanto se necesitó utilizar un concentrador activo de 8 puertos de conexión al cual todas las computadoras estarán conectadas por medio de cable conocido como UTP categoría cinco, (parecido al

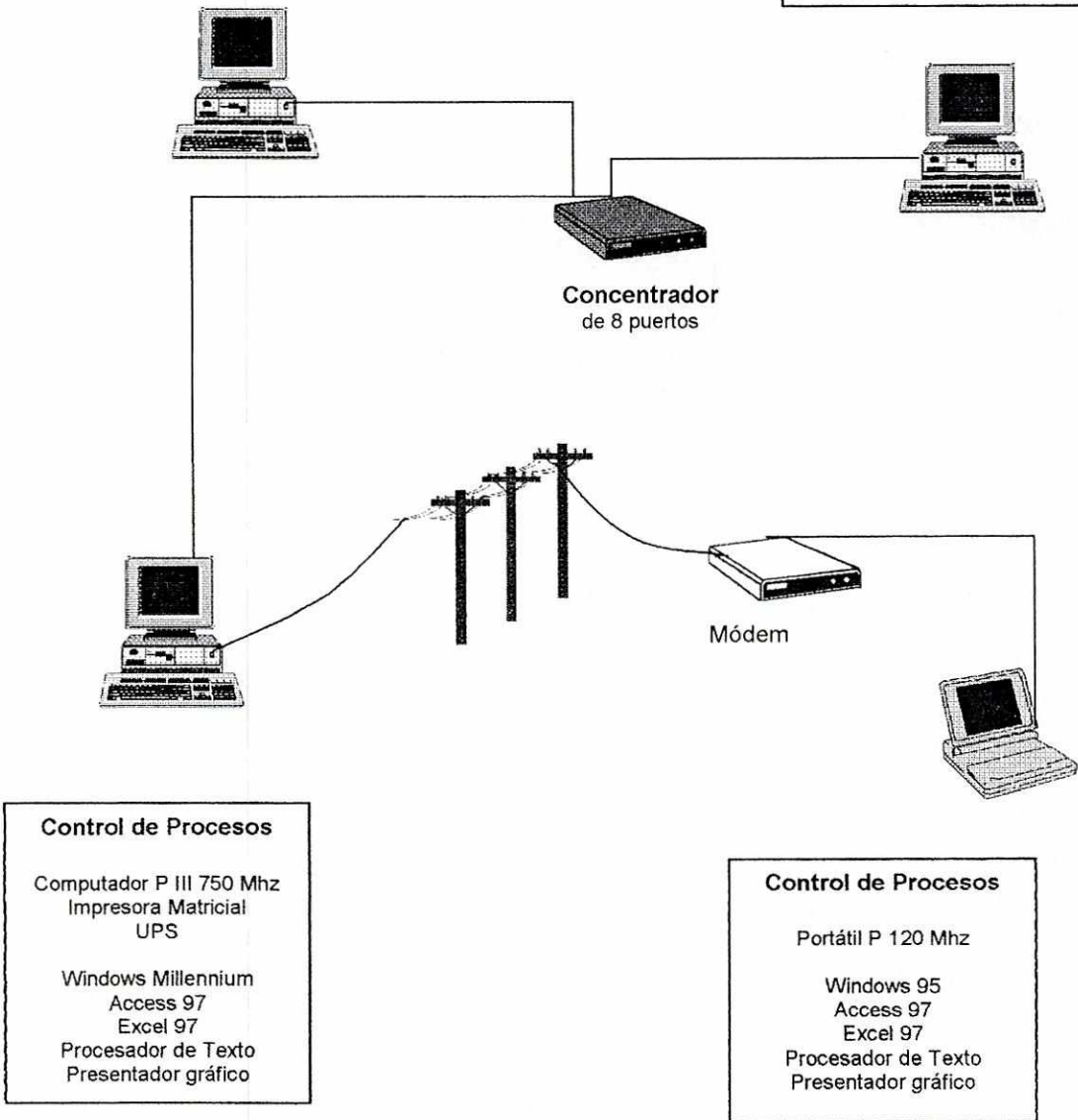
cable telefónico utilizado en el interior de una vivienda). El siguiente esquemático ilustra lo anteriormente descrito.

Recepción
Computador P II 500 Mhz
Impresora de Inyección
UPS

Windows 98
Access 97
Excel 97
Procesador de texto

Dirección de Análisis
Computador P III 750 Mhz
Impresora de Inyección
UPS

Windows Millennium
SPSS
Access 97
Excel 97
Procesador de texto
Presentador gráfico



Control de Procesos
Computador P III 750 Mhz
Impresora Matricial
UPS

Windows Millennium
Access 97
Excel 97
Procesador de Texto
Presentador gráfico

Control de Procesos
Portátil P 120 Mhz

Windows 95
Access 97
Excel 97
Procesador de Texto
Presentador gráfico

VI. OBSERVACIONES FINALES

- Al igual que se expone en la sección inicial de este modelo, el ser humano utiliza herramientas desde sus orígenes; no importando la época que se vive, siempre habrá un desarrollo tecnológico para ese momento particular y hará uso de esa técnica que pone las herramientas a su disposición.
- Es importante que antes de utilizar esas herramientas, se dé un paso previo que es el de la obtención de la información. Qué mejor que estar al día tecnológicamente y aumentar la eficiencia. Sin embargo si se hace, un análisis exhaustivo, un planeamiento estratégico, un diseño de actividades e implementación y pruebas, se estará utilizando, no sólo la herramienta apropiada, sino aprovechándola con su mejor punto de apoyo para lograr los mejores beneficios de la misma.
- Cuando se pretende utilizar una herramienta aplicada a los diferentes campos profesionales, en este caso de informática y computación, es importante que se utilicen los conocimientos de un profesional en el área, a manera de consulta inicial, para ahorrarse los pasos de la experimentación que ya alguien más los ha vivido. No necesariamente, en algunos casos se le contratará para formar parte del personal de planta de la empresa, pero sí para trazar los lineamientos de su especialidad.
- Debiera escogerse no sólo un profesional en el área de cómputo, sino que esté bien relacionado con el tipo de investigación que se necesita en ese caso particular, esto facilitará el éxito del estudio y propiciará la toma correcta de decisiones a los responsables de la ejecución de proyectos y/o actividades que respondan a los objetivos de las Empresas en estudio.
- Es necesario recabar y dar toda la información existente, relacionada con la aplicación, al consultor, para diseñar la propuesta correcta para evitar el doble manejo de información y así economizar tiempo y recursos económicos.
- Luego de tener claramente definidos los requerimientos para implementar la solución para la empresa, se puede comprar el hardware que responda a dichos requerimientos específicos. Sin olvidar que aunque se dé inicio con un ambiente pequeño o mediano, éste debe contemplar que deje interfaces abiertas y flexibles para el crecimiento.

- Debe considerarse hacer innumerables pruebas, no sólo del hardware, sino especialmente del software aplicado, ej. En este caso el programa SPSS y ver que los resultados sean satisfactorios y complacientes al usuario final.
- Todas las observaciones descritas anteriormente, deben llevarse a cabo para alcanzar los objetivos empresariales pero; ¿cómo pueden hacerse todas y cada una de ellas efectivas? A través de una capacitación constante sobre los avances tecnológicos, así como los cambios y reformas administrativas para el manejo eficiente de la información.

VII. BIBLIOGRAFIA

- Arkin, Herbert, Colton Raymond: Métodos Estadísticos, México, D.F., Editorial Continental, 1-68 pp.
- Chacón, Moscoso S.; Lopez, Ruiz J.: La Metodología de la Evaluación de Programas de Intervención: Una Aplicación en Centros Asistenciales Infantiles. Apuntes de Psicología: No. 37. 41-60 pp.
- Digital, Equipment Corporation: Introduction to Basic. 1978 Maynard, Massachusetts. 12 pp.
- Elorza, Haroldo: Estadística para las Ciencias Sociales y del Comportamiento, México, 1999 D.F. Editorial Oxford, Segunda Edición. 127 pp.
- Fitz – Gilbou, Carol T.: Como Diseñar Evaluación de un Programa, Universidad de New York 1981 York 1981. 185 pp.
- García Ferrando Manuel: El Análisis de la Realidad Social, Métodos y Técnicas de Evaluación, Madrid, España. Segunda Edición, Editorial Alianza..
- Glossbrenner, Alfred: Personal Computer Communications, New York 1983 Edit. St. Martin's Pres 308 pp.
- Hopkins, Kenneth D., Hopkins B.R.: Estadística Básica, México, D.F., Editorial Prentice Hall, 35-99 pp.
- House, Ernest R.: Enfoques de Evaluación, Universidad de New York 1981
- Jordan, Larry: Sistem Integration for the IBM PS/2 and PC, New York,, 1990 Edit.Brady, 203 pp.
- Mendenhall, Scheaffer: Elementos de Muestreo, México, D.F. Editorial Iberoamérica 1987 19-37 pp.
- Novell, Personal NetWare: Personal NetWare User Guide. Provo, UT. U.S.A. 1993 Edit. Novell Inc., 1-3 pp.
- Page, William G.: Using Oracle 8/8i; Indianapolis, Indiana. U.S.A. 1999 Edit. Macmillan Computer Publishing, 57 pp.
- Parsaye, Kamran, Chignell, Mark, Khoshafian, Setrag Wong, Harry: Intelligent Databases 1989 New York.Edit. John Wiley & Sons, Inc. v pp.
- Pinter, Les; Pinter, John: Visual FoxPro Manual de Programación 3.0 y 5.0 1997 México, D.F.Editorial Mc Graw Hill. 298 pp.
- Prague, Cary N.; Amo, William C.;Foxall James D.: Acces 97 Secrets, Foster City, C.A. 1997 Edit. IDG Books Worldwide, 35 pp.

Senn, James A.: Análisis y Diseño de Sistemas de Información. México, D.F.
1998 Editorial Mc Graw Hill Segunda Edición. 117pp.

Spiegel, Murray R.; Stephens, Larry J.: Statistics. New York. Editorial McGraw
1998 Hill Company Third Edition. 84 pp.

Spiegel, Murray R.: Estadística. Colección Schaum, México, D.F. Segunda Edición
1998 65, 83, 289, 119, 61 pp.

Stevenson, William: Estadística para Administración y Economía
1978 México, D.F. Editorial Harla 23-35 pp.

Tomayko, James Dr.: Management of Software Development. Jet@cs.cmu.edu
1995 Readings .Wean Fall 4524. 23pp.

Taylor, Grant: Linux Complete. Alameda, C.A.Edit. Sybex Inc.
1999 xxxv pp.

Turk, Thomas A.: Planning and Desingning the data base Environment. New York.
1985 Editorial Van Nostrand Reinhold Company. 120pp.

VIII. GLOSARIO

Nota: Las definiciones contenidas en este glosario pertenecen generalmente al campo de Estadística, Comunicaciones y Sistemas de Computación.

algoritmos genéticos: Técnicas de optimización que usan procesos tales como combinación genética, mutación y selección natural en un diseño basado en los conceptos de evolución natural.

amplitud de variación: (rango) Es generalmente la medida más sencilla y fácil de comprender y calcular. Se concentra en el número mayor y menor del grupo. Dicha medida se puede calcular a través de la diferencia entre los valores mayor y menor.

análisis de series de tiempo (time-series): Análisis de una secuencia de medidas hechas a intervalos específicos. El tiempo es usualmente la dimensión dominante de los datos.

análisis prospectivo de datos: Análisis de datos que predice futuras tendencias, comportamientos o eventos basado en datos históricos.

análisis exploratorio de datos: Uso de técnicas estadísticas tanto gráficas como descriptivas para aprender acerca de la estructura de un conjunto de datos.

análisis retrospectivo de datos: Análisis de datos que provee una visión de las tendencias comportamientos o eventos basado en datos históricos.

árbol de decisión: Estructura en forma de árbol que representa un conjunto de decisiones. Estas decisiones generan reglas para la *clasificación* de un conjunto de datos. Ver *CART* y *CHAID*.

asimetría: (sesgo) El grado de asimetría de una distribución, es decir, cuánto se aparta de la simetría.

atributo: Cualidad o característica que se atribuye a los elementos de una población y permite clasificar (formar grupos).

base de datos: La forma organizada para almacenar *data*, a través de una o más tablas, las cuales pueden llevar consigo relaciones directas o indirectas entre sí.

base de datos multidimensional: Base de datos diseñada para procesamiento analítico on-line (*OLAP*). Estructurada como un hipercubo con un ejemplo por dimensión.

baudio: Una unidad para medir la velocidad de transmisión de datos. Técnicamente se refiere al número de veces que la línea de comunicaciones cambia de estado cada segundo. Baudio y bit no son medidas idénticas

bit: Es el acrónimo para "dígito binario". Es la unida más pequeña de información en el mundo de computación. Dependiendo el tipo de computador, 8 bits juntos hacen un "byte"

byte: Se le conoce en la tecnología de sistemas de cómputo, al conjunto de 8 bits o más bits, dependiendo el tipo de computador.

cable coaxial: Tipo de cable utilizado en comunicaciones de datos o comunicaciones de radio que utiliza un cable de dos polaridades, uno medio aislado para positivo y un externo de malla generalmente para negativo, con un aislamiento exterior. La característica primordial es que posea una resistencia de 52 ohmios, con un diámetro de 5 milímetros en el tipo delgado o de 11 milímetros en el de tipo grueso.

cart arboles de clasificación y regresión: Una técnica de *árbol de decisión* usada para la *clasificación* de un conjunto da datos. Provee un conjunto de reglas que se pueden aplicar a un nuevo (sin clasificar) conjunto de datos para predecir cuáles registros darán un cierto resultado. Segmenta un conjunto de datos creando 2 divisiones. Requiere menos preparación de datos que *CHAID*.

chaid detección de interacción automática de chi cuadrado: Una técnica de *árbol de decisión* usada para la *clasificación* de un conjunto da datos. Provee un conjunto de reglas que se pueden aplicar a un nuevo (sin clasificar) conjunto de datos para predecir cuáles registros darán un cierto resultado. Segmenta un conjunto de datos utilizando tests de chi cuadrado para crear múltiples divisiones. Antecede, y requiere más preparación de datos, que *CART*.

censo: Comprende el examen de todos y cada uno los elementos de un universo o población.

clasificación: Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a variable(s) específica(s) las cuales se están tratando de predecir. Por ejemplo, un problema típico de clasificación es el de dividir una base de datos de compañías en grupos que son lo más homogéneos posibles con respecto a variables como "posibilidades de crédito" con valores tales como "Bueno" y "Malo".

clase: Sub población que posee cierto atributo o característica, al resumir grandes colecciones de datos es útil distribuirlos en clases o categorías y determinar el número de elementos que pertenecen a cada clase, llamado frecuencia de clase.

clustering (agrupamiento): Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a todas las variables disponibles.

computadoras con multiprocesadores: Una computadora que incluye múltiples procesadores conectados por una red. Ver *procesamiento paralelo*.

comunicación asincrónoma: Comunicación de datos en la variedad de principio a fin. Cada carácter es transmitido como una unidad discreta con su propio bit de inicio y uno o más bits de parada.

comunicación sincrónoma: *La técnica de comunicación de datos en la cual los bits son transmitidos y recibidos a una velocidad fija. Se utiliza para transmitir grandes blocks de datos en líneas especiales de comunicación. Es mucho más compleja que la comunicación asincrónoma. Tiene muy poca aplicación actualmente.*

concentrador: *Dispositivo que actúa como central para recibir físicamente las conexiones de una o más computadoras en red. También recibe y transmite la información que pasa por sus componentes electrónicos.*

correlación: Es la medida de concordancia entre dos variables. En las ciencias sociales, se usan medidas estadísticas para formular predicciones. Estos pronósticos se basan en la relación entre dos variables. Las medidas de correlación son indicadores o medidas de vínculos o relaciones.

curtosis: La curtosis mide cuán escarpada o aguda es la gráfica (polígono de frecuencias) en una distribución, en general. Una medida de la curtosis utiliza el cuarto momento respecto de la media en forma adimensional.

data: La forma plural de *datum* pero comúnmente utilizado en singular; cualquier elemento de información que puede ser procesado por una computadora.

data cleansing: Proceso de asegurar que todos los valores en un conjunto de datos sean consistentes y correctamente registrados.

data mining: La extracción de información predecible escondida en grandes bases de datos.

data warehouse: Sistema para el almacenamiento y distribución de cantidades masivas de datos.

datos anormales: Datos que resultan de errores (por ejemplo: errores en el tipeado durante la carga) o que representan eventos inusuales.

desviación absoluta media: Mide la desviación promedio de valores, con respecto a la media del grupo, sin tomar en cuenta el signo de la desviación.

desviación estándar: Se define como la raíz cuadrada de la varianza.

dimensión: En una base de datos relacional o plana, cada campo en un registro representa una dimensión. En una *base de datos multidimensional*, una dimensión es un conjunto de entidades similares; por ejemplo: una base de datos multidimensional de ventas podría incluir las dimensiones Producto, Tiempo y Ciudad.

distribuciones de frecuencias: Es un arreglo o agrupamiento de datos en clases que muestra, el número o porcentaje (es decir la frecuencia), de observaciones de cada una de ellas. Una distribución de frecuencia se puede presentar en forma tabular o gráfica. El número o porcentaje en una clase se denomina frecuencia de clase.

error estándar de estimación: Se llama así al error de medida que existe cuando la relación entre variables, no es perfecta. Este se obtiene de la diferencia entre la puntuación predicha y la observada. En muestreo es el valor que se suma y resta a la estimación puntual para obtener el intervalo de confianza.

escala de intervalos: Esta es más refinada, tanto que números consecutivos marcan intervalos iguales, es decir cantidades iguales de la variable que se mide. El cero es arbitrario, es decir que el cero no representa.

escalas de medición: La medición incluye la evaluación, es un proceso mediante el cual las cosas se diferencian.

escala nominal: Es la forma más rudimentaria de medir. Es el proceso de agrupar unidades (objetos, personas etc.), en categorías basándose en uno o más atributos o propiedades observadas. En las escalas nominales o categóricas, los números asignados, definen cada grupo distinto y sirven especialmente como etiquetas, nombres o códigos de identidad.

escala ordinal: Se logra cuando las observaciones pueden colocarse en un orden o una jerarquía con respecto a la característica que se evalúa. La magnitud de los números representa el orden del rango del atributo observado.

escala de razón o de proporción: Tiene las propiedades de la escala ordinal y de intervalo, pero además el cero es absoluto, representa la ausencia de la característica en cuestión; en consecuencia los números pueden compararse como proporciones.

estadígrafo: Es una característica representativa de la muestra, es utilizado para designar cualquier función de datos empíricos utilizados con fines descriptivos.

estadígrafos de tendencia central: Conllevan información respecto al valor promedio de un conjunto de valores. Las medidas más frecuentes son: la moda, la media y la mediana.

estadígrafos de variabilidad: Estas medidas indican si los valores están relativamente cercanos uno del otro o si se encuentran dispersos. La dispersión aumenta a medida que se incrementa el valor de la medida. Excepto la amplitud de variación, todas las medidas de dispersión toman como punto de referencia a la media.

estadística descriptiva: Es la rama de la Estadística que organiza, tabula, simplifica, presenta y describe en términos generales, conjuntos de datos; éstos datos pueden ser de variables cuantitativas o cualitativas.

estadística inferencial: Consiste en el análisis e interpretación de una muestra de datos, proporciona métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones. Su propósito principal es estimar los atributos de la población a partir de solo una muestra de casos.

estadística probabilística: Es de gran utilidad para el análisis de situaciones en las que interviene el azar. Esta es parte de la estadística descriptiva.

experimento: Es una actividad planificada cuyos resultados producen un conjunto de datos.

frecuencia acumulada: Está diseñada para indicar el número o porcentaje de elementos que son menores o iguales que cierto valor específico.

full duplex: Un acondicionamiento de comunicaciones que permite enviar y recibir información simultáneamente, por ejemplo la forma de comunicación cuando se habla por teléfono.

half duplex: Un acondicionamiento de comunicaciones que permite que los datos puedan viajar en una sola dirección a la vez. Por ejemplo en la comunicaciones por radio transmisores y receptores.

hardware: El equipo o sus partes físicas que componen un sistema de computación.

mainframes: La más grande de las variedades que hay entre computadores. Por ejemplo, en orden descendente estarían las minicomputadoras y las microcomputadoras.

media: Se usa con datos de intervalo o de razón. Se asocia a una distribución simétrica normal.

media aritmética: Es el promedio aritmético de un conjunto de observaciones (x). Se obtiene como el cociente de dividir la sumatoria de los valores de la variable entre el tamaño de la población.

media armónica: La media armónica H de un conjunto de números, es el recíproco de la media aritmética de los recíprocos de esos números.

media cuadrática: Se define como la raíz cuadrada del promedio aritmético de los cuadrados de los valores de la variable. Entre el tamaño de la población.

media geométrica: La media geométrica G de un conjunto de números positivos es la raíz N-ésima del producto de esos números.

mediana: Es el valor central en un conjunto de valores ordenados: el punto abajo - arriba del cual cae un número igual de observaciones, la mediana es el percentil 50 (P50) o el segundo cuartil (Q2). La mediana puede calcularse desde una distribución que pueda ordenarse.

media ponderada: Es el promedio aritmético que al sumar los valores de la variable considerada su "peso frecuencial" previo al cociente entre el tamaño de la población.

memoria: Un dispositivo en el cual los datos pueden ser almacenados y también pueden ser extraídos; también se le conoce como memoria interna de almacenaje y core.

moda: Es la observación que ocurre con mas frecuencia, es el valor que mas se repite en la población. Desde el punto de vista puramente descriptivo, la moda es indicativa del valor "típico", en términos relativos es el valor que adquiere el mayor porcentaje frecuencial.

modelo analítico: Una estructura y proceso para analizar un conjunto de datos. Por ejemplo, un *árbol de decisión* es un modelo para la *clasificación* de un conjunto de datos.

modelo lineal: Un *modelo analítico* que asume relaciones lineales entre una variable seleccionada (dependiente) y sus predictores (variables independientes).

modelo no lineal: Un *modelo analítico* que no asume una relación lineal en los coeficientes de las variables que son estudiadas.

modelo predictivo: Estructura y proceso para predecir valores de variables especificadas en un conjunto de datos.

módem: Un acrónimo para "modulador/demodulador" Este es el dispositivo que traduce las señales entrantes desde un computador, a una forma que puedan ser transmitidas por las líneas telefónicas. También puede traducir señales entrantes en una forma que dos

computadoras puedan entenderse. Dos módems, uno para cada computadora son necesarios para cualquier comunicación de datos a través de las líneas telefónicas.

muestra: Puede definirse como una parte de la población o universo que sea representativa, significativa y confiable.

navegación de datos: Proceso de visualizar diferentes dimensiones y niveles de una *base de datos multidimensional*. Ver *OLAP*.

nodo: Un sub sistema de operación completo, que forma parte de un cadena o de un grupo, el cual está enlazado con otros sistemas afines. Ej. Una estación de trabajo de una red de computadoras.

olap procesamiento analítico on-line (On Line Analytic processing): Se refiere a aplicaciones de bases de datos orientadas a array que permite a los usuarios ver, navegar, manipular y analizar *bases de datos multidimensionales*.

outlier: Un ítem de datos cuyo valor cae fuera de los límites que encierran a la mayoría del resto de los valores correspondientes de la muestra. Puede indicar *datos anormales*. Deberían ser examinados detenidamente; pueden dar importante información.

procesamiento paralelo: Uso coordinado de múltiples procesadores para realizar tareas computacionales. El procesamiento paralelo puede ocurrir en una computadora con múltiples procesadores o en una red de estaciones de trabajo o PCs.

parámetro: Es una característica o atributo medible de una población completa.

población: Es un conjunto de elementos, una población de valores se puede definir como la mayor colección de valores para la variable aleatoria, los cuales son de interés en un momento particular.

procedimiento estadístico: Los métodos principales para organizar datos estadísticos comprenden el ordenamiento de elementos en subconjuntos que presenten cualidades semejantes. Los datos agrupados se pueden resumir gráficamente, o en tablas y mediante el uso de medidas numéricas. El nombre que reciben los datos ordenados en grupos o categorías es el de distribución de frecuencias.

programa: La secuencia completa de instrucciones que siguen un orden lógico, datos y rutinas necesarios para la solución de un problema. En términos generalizados se le conoce como el conjunto de instrucciones creados en un lenguaje de computadora, para producir un resultado.

protocolo: Otro nombre para definir "el acondicionamiento de comunicaciones" para que dos equipos electrónicos puedan entenderse entre sí. Se dice también en términos comunes, que las dos equipos "hablan en mismo idioma".

prototipo: Es un sistema que funciona - no sólo una idea en el papel - desarrollado con la finalidad de probar ideas y suposiciones relacionadas con el nuevo sistema.

puntero: Otro nombre para designar al selector de un registro específico cuando se hace una búsqueda directa en una tabla que forma parte de la base de datos.

regresión lineal simple: Es un procedimiento estadístico para predecir el desempeño en variables de criterio de una o más variables predictoras.

sdlc: Una estrategia para el desarrollo de sistemas. Por su siglas en Inglés: System Development Life Cycle (Ciclo de Vida para Desarrollo de Sistemas)

sistema de computación: La combinación de hardware y software que desarrolla operaciones de procesamiento específico; una colección de componentes que forman una unidad funcional.

servidor: El computador central que actúa como central de proceso para varias estaciones de trabajo. Por lo general contendrá los archivos más importantes que deben de ser compartidos con los usuarios.

software: Todos los programas, procedimientos, reglas e información periférica, asociada con la operación de un sistema de computación.

software de comunicaciones: El programa de computadora que contiene las instrucciones que necesita su equipo para poderse comunicar con un determinado computador.

tabla: Una colección de datos almacenados para referencia fácil; data almacenada en un arreglo de filas y columnas.

topología: La forma de estructuración de los componentes de un sistema de red. Los

componentes incluyen: tipo de cable, distribución del cableado, rutas a seguir, distribución de los computadores, estaciones de trabajo y/o servidores.

universo: Conjunto total de todos los datos existentes de personas, objetos o cosas. Se denota como población finita de N unidades.

variable: Una letra, carácter, símbolo, que puede asumir cualquier valor de un grupo de valores dados; la representación simbólica en el almacenaje, cuyo valor puede cambiar durante un programa.

variables cuantitativas: Ya sean continuas o discretas, son aquellas que pueden medirse en forma usual. Las mediciones hechas sobre este tipo de variables, conllevan información respecto a cantidad. Es decir, datos que pueden expresarse con números y explican "cantidades".

variables cualitativas o categóricas: Se refiere a una cualidad, contienen información respecto a los atributos. Son difíciles de cuantificar, suelen codificarse.

variable dependiente: Es la que el investigador no puede manipular, es la variable que depende de otras, es decir es la variable resultante de una o más variables independientes o predictores. Se ubica en el eje vertical usualmente.

variable estadística: Es una característica de cada elemento individual de una población o muestra, cuyo valor no es constante.

variable independiente: Es la variable que el investigador puede manipular, así como darle valores supuestos, usualmente se ubica en el eje horizontal. La variable independiente, sirve para predecir una relación entre *esta y la dependiente*.

varianza: Es un estadígrafo de variabilidad o de dispersión. Se obtiene sumando los cuadrados de los desvíos de los valores de la variable respecto a su media aritmética y dividiendo dicha suma entre el tamaño de la población (varianza sesgada). Si se divide entre el tamaño poblacional menos uno ($N-1$), se transforma en varianza "insesgada" (muestreo).

ANEXO



MENU PRINCIPAL

VALIDACION DE INSTRUMENTOS DE EVALUACION

CAPTURA DE DATOS PARA LAS DIFERENTES BASES RELACIONALES

Instrumento para Características Regionales

Instrumento para Situación Habitacional

Instrumento para Situación Familiar

Departamentos

Municipios

Aldeas

Caseños

Encuestadores

Proyectos

Tipos de Proyectos



Registro: 1 de 1

Vista Formulario



Encuesta1

INGRESO DE DATOS PARA EL INSTRUMENTO PARA CARACTERISTICAS

| | |
|--------------------|----------------------|
| CORRELATIVO | |
| FECHA | 30/04/99 |
| Nombre_localidad | ALDEA CIENAGA GRANDE |
| Codigo_localidad | 00008 |
| Departamento | CHIMALTENA |
| Municipio | CHIMALTENANGO |
| Aldea | CIENAGA GRANDE |
| Caserio | |
| Cantidad_casas | 9207 |
| Cantidad_personas | |
| Encuestador_codigo | 009 |



Registros: 31 de 66

Vista Formulario



ENCUESTA 3 DET

CAPTURA DE DATOS DE INSTRUMENTO PARA SITUACION FAMILIAR

Pantalla de Detalles

| | |
|------------------|-----|
| CODIGO ENCUESTA | 144 |
| NOMBRE ADICIONAL | |
| PARENTESCO | 2 |
| EDAD | 39 |
| GENERO | F |
| ESCOLARIDAD | AL |
| GRADO | P |
| SOSTIENE | S |
| RELIGION | 2 |



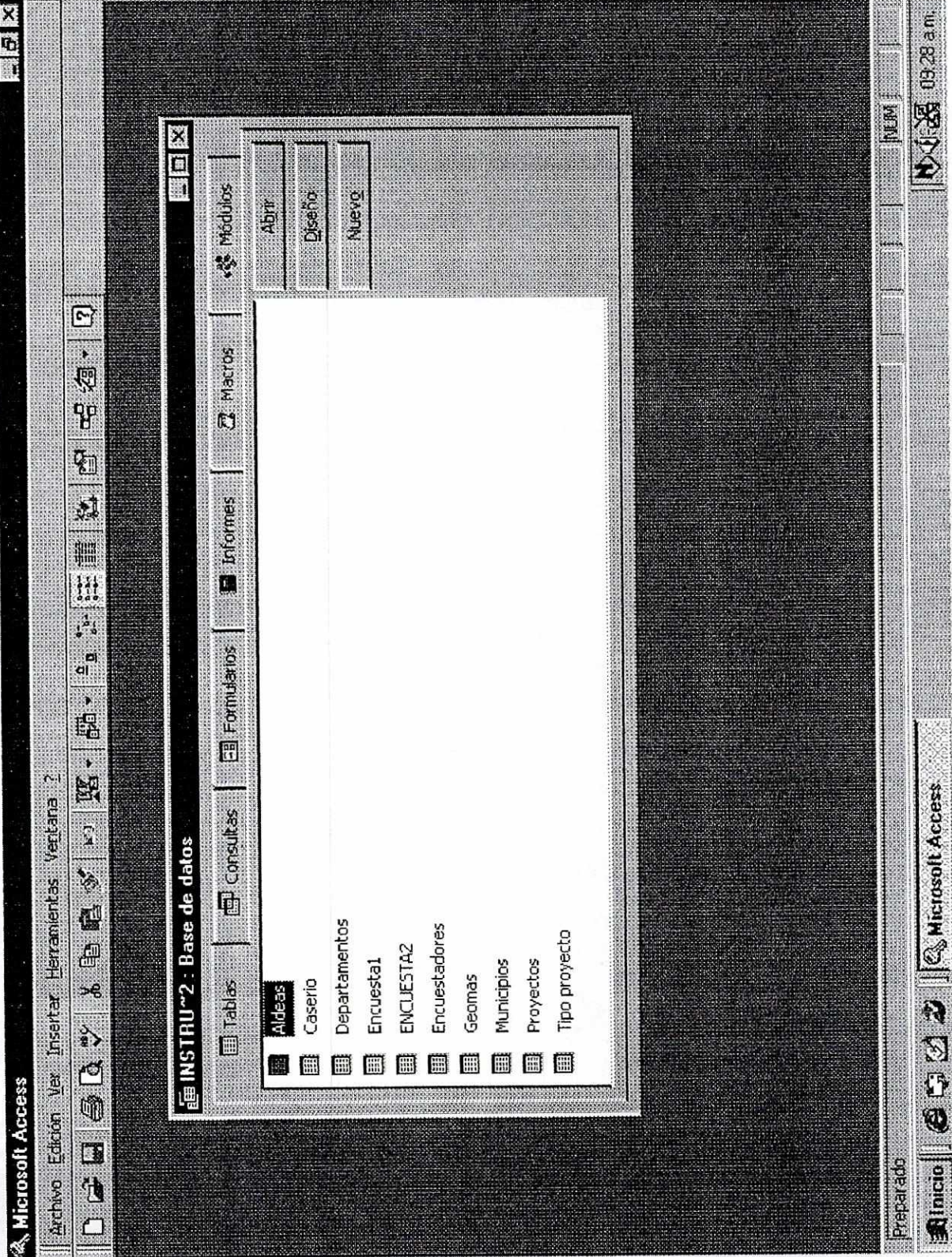
Registros: 1 de 539

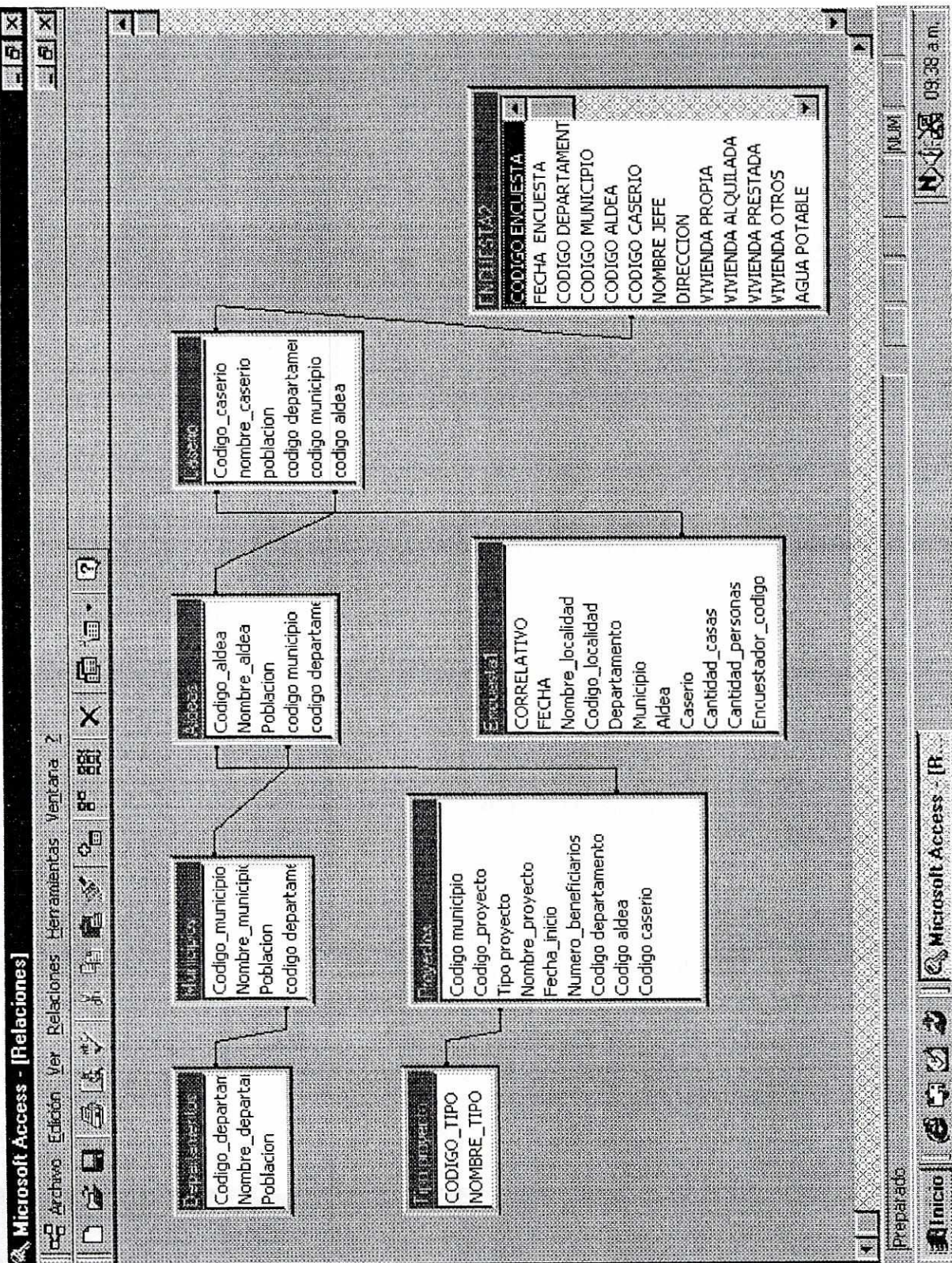
Vista Formulario



Proyectos : Tabla

| | Codigo_munic | Codigo_proye | Tipo_proyecto | Nombre_proyecto | Fecha_inicio | Numero_bene | Con |
|---|--------------|--------------|---------------|----------------------|--------------|-------------|-----|
| ▶ | 01 | 0501797 | 03 | ARTESANIAS | 01/01/98 | 50 | 01 |
| | 01 | 0501597 | 10 | CONSTRUC. CARRETERAS | 01/01/98 | 50 | 01 |
| | 01 | 0501497 | 07 | VIVERO FORESTAL | 01/01/98 | 50 | 01 |
| | 01 | 0501197 | 01 | ALFABETIZACION | 15/01/98 | 50 | 01 |
| | 01 | 0501897 | 08 | CRIANZA DE POLLOS | 01/01/98 | 50 | 01 |
| | 01 | 0501697 | 07 | VIVERO FORESTAL | 01/01/98 | 25 | 01 |
| | 01 | 0501397 | 01 | ALFABETIZACION | 15/01/98 | 50 | 01 |
| | 01 | 0501297 | 06 | VIVERO COMUNAL | 15/01/98 | 50 | 01 |
| | 01 | 0501097 | 06 | VIVERO COMUNAL | 01/01/98 | 50 | 01 |
| | 01 | 0500997 | 06 | VIVERO COMUNAL | 01/01/98 | 25 | 01 |
| | 01 | 0501497 | 01 | ALFABETIZACION | 01/01/98 | 50 | 01 |
| | 01 | 0501497 | 08 | CRIANZA DE POLLOS | 01/01/98 | 50 | 01 |
| | 01 | 0501897 | 09 | CRIANZA DE CERDOS | 01/01/98 | 50 | 01 |
| | 01 | 0501697 | 06 | VIVERO COMUNAL | 01/01/98 | 25 | 01 |
| | 01 | 0501397 | 03 | ARTESANIAS | 15/01/98 | 50 | 01 |
| | 01 | 0501297 | 01 | ALFABETIZACION | 15/01/98 | 50 | 01 |
| | 01 | 0501097 | 01 | ALFABETIZACION | 01/01/98 | 50 | 01 |
| | 01 | 0500997 | 01 | ALFABETIZACION | 01/01/98 | 25 | 01 |
| | 01 | 0508597 | 12 | CULTIVO DE GRANADIA | 15/01/98 | 25 | 01 |
| | 03 | 0500497 | 03 | ARTESANIAS | 05/10/97 | 253 | 01 |
| | 03 | 0500397 | 06 | VIVERO COMUNAL | 28/11/97 | 25 | 01 |





**Proyecto de Infraestructura
Tablas - Empresa de Investigación
Reporte de cubo con tabla pivotada**

Reporte Olap de Contraste

| | TRANSPORTE CARRO PROPIO | | | | | | | | | |
|----------------|-------------------------|----|-------|----|-------|-----------------|--|-------|--|-------|
| | No | | | | | Col % | | | | |
| | Count | | | | | VIVIENDA PROPIA | | | | |
| | VIVIENDA PROPIA | | Total | | | VIVIENDA PROPIA | | Total | | |
| | No | Si | | | No | Si | | | | Total |
| CODIGO | 1 | 14 | | 15 | 20.0% | 31.1% | | | | 30.0% |
| ALDEA | 1 | 10 | | 11 | 20.0% | 22.2% | | | | 22.0% |
| | 3 | 1 | | 1 | 60.0% | 2.2% | | | | 2.0% |
| Hacienda Maria | | 20 | | 23 | | 44.4% | | | | 46.0% |

Reporte Olap de Contraste

| | TRANSPORTE CARRO PROPIO | | | | | | | | | |
|----------------|-------------------------|-------|-------|-------|------|-----------------|--|-------|--|-------|
| | No | | | | | Table % | | | | |
| | Layer % | | | | | VIVIENDA PROPIA | | | | |
| | VIVIENDA PROPIA | | Total | | | VIVIENDA PROPIA | | Total | | |
| | No | Si | | | No | Si | | | | Total |
| CODIGO | 20.0% | 31.1% | | 30.0% | 2.0% | 28.0% | | | | 30.0% |
| ALDEA | 20.0% | 22.2% | | 22.0% | 2.0% | 20.0% | | | | 22.0% |
| | 60.0% | 2.2% | | 2.0% | 6.0% | 2.0% | | | | 2.0% |
| Hacienda Maria | | 44.4% | | 46.0% | | 40.0% | | | | 46.0% |

Reporte Olap de Contraste

| | TRANSPORTE MOTO PROPIA | | | TRANSPORTE BICICLETA | | |
|----------------|------------------------|----|-------|----------------------|----|-------|
| | VIVIENDA PROPIA | | Total | VIVIENDA PROPIA | | Total |
| | No | Sí | | No | Sí | |
| CODIGO | 1 | 14 | 15 | 1 | 14 | 15 |
| ALDEA | 1 | 10 | 11 | 1 | 10 | 11 |
| Saquitacaj | 1 | 1 | 1 | 1 | 1 | 1 |
| Hacienda María | 3 | 20 | 23 | 3 | 20 | 23 |

Reporte Olap de Contraste

| | TRANSPORTE CAMION | | | TRANSPORTE CAMIONETA | | |
|----------------|-------------------|----|-------|----------------------|----|-------|
| | VIVIENDA PROPIA | | Total | VIVIENDA PROPIA | | Total |
| | No | Sí | | No | Sí | |
| CODIGO | 1 | 14 | 15 | 4 | 4 | 4 |
| ALDEA | 1 | 10 | 11 | | | |
| Saquitacaj | 1 | 1 | 1 | 1 | 1 | 1 |
| Hacienda María | 3 | 20 | 23 | 5 | 5 | 5 |

Reporte Olap de Contraste

| | TRANSPORTE CAMIONETA | | | | TRANSPORTE CABALLO | | | |
|--------|----------------------|----|-------|-----------------|--------------------|-------|--|--|
| | Sí | | | | No | | | |
| | VIVIENDA PROPIA | | Total | VIVIENDA PROPIA | | Total | | |
| | No | Sí | | No | Sí | | | |
| CODIGO | 1 | 10 | 11 | 1 | 14 | 15 | | |
| ALDEA | 1 | 10 | 11 | 1 | 10 | 11 | | |
| | 3 | 15 | 18 | 3 | 1 | 1 | | |
| | | | | | 15 | 18 | | |

Reporte Olap de Contraste

| | TRANSPORTE CABALLO | | | | TRANSPORTES OTROS | | | | |
|--------|--------------------|----|-------|-----------------|-------------------|-------|-----------------|----|-------|
| | Sí | | | | No | | | | |
| | VIVIENDA PROPIA | | Total | VIVIENDA PROPIA | | Total | VIVIENDA PROPIA | | Total |
| | No | Sí | | No | Sí | | No | Sí | |
| CODIGO | | | | 1 | 10 | 11 | 4 | 4 | |
| ALDEA | | | | 1 | 10 | 11 | 1 | 1 | |
| | | 5 | 5 | 3 | 20 | 23 | | | |

Reporte de cubo TransVivieElectActiv Infra - SPSS Viewer

File Edit View Insert Evol Format Analyze Graphs Utilities Window Help

Output
 Tables
 Title
 Notes
 Reporte de Cubo

Proyecto Infraestructura Tablas pivotadas - Empresa de Investigación

Reporte de Cubo Interactivo

| | | | | | |
|----------------------|--------------------------|-------------------|----------------------|--------------------|----|
| Layers | VIVIENDA PROPIA SI Count | | | | |
| Rows | ELECTRODOMESTICOS SI | | | | |
| TRANSPORTE CARR PROP | ELECTRODOMESTICOS SI | | | | |
| No. | Total | | | | |
| | GRABADORAS .00 | | | | |
| | GRABADORAS 1.00 | | | | |
| | GRABADORAS 2.00 | | | | |
| | Total | | | | |
| | EQUIPO DE SONIDO .00 | | | | |
| | Total | | | | |
| | | TRANSPORTE CAMION | TRANSPORTE CAMIONETA | TRANSPORTE CABALLO | |
| | | No | Si | No | Si |
| 7 | 17 | 4 | 13 | 17 | 17 |

SPSS Processor is ready

Encuesta2 - SPSS Data Editor

Reporte de cubo Tra...

Inicio

H: 125 W: 554 pt

07:08 a.m.

Proyecto de Reforestación y Medio Ambiente
Tabla estadística - Empresa de Investigación

Reporte de Cubo - Empresa de Investigación

| | BASURA PATIO CASA | | | LEJOS DE CASA | | DRENAJE RIO | |
|------------------|-------------------|-------|--------|---------------|----|-------------|----|
| | No | | Row % | No | Sí | | |
| | Count | Col % | | | | | |
| CODIGO ALDEA | Pameya | 15 | 30.0% | 100.0% | 13 | 2 | 15 |
| | Palama | 11 | 22.0% | 100.0% | 9 | 2 | 11 |
| | Saquitacaj | 1 | 2.0% | 100.0% | 1 | | 1 |
| | Hacienda Maria | 23 | 46.0% | 100.0% | 10 | 13 | 23 |
| CODIGO MUNICIPIO | San José Poaquil | 50 | 100.0% | 100.0% | 33 | 17 | 50 |

Reporte de Cubo - Empresa de Investigación

| | LA ENTIERRA | | TIRA EN CUALQUIER LUGAR | LA QUEMA | | |
|------------------|------------------|----|-------------------------|----------|----|----|
| | No | Sí | | No | Sí | |
| | | | | | | 15 |
| CODIGO ALDEA | Pameya | 9 | 2 | 11 | 6 | 5 |
| | Palama | 1 | | 1 | | 1 |
| | Saquitacaj | 19 | 4 | 23 | 21 | 2 |
| | Hacienda Maria | 44 | 6 | 50 | 33 | 17 |
| CODIGO MUNICIPIO | San José Poaquil | | | | | |

Reporte de Cubo - Empresa de Investigación

| | EN CAMION | BASURA OTROS | |
|------------------|-----------|--------------|----|
| | | No | Si |
| CODIGO ALDEA | | | |
| Pameya | 15 | 8 | 7 |
| Palama | 11 | 9 | 2 |
| Saquitacaj | 1 | 1 | |
| Hacienda María | 23 | 19 | 4 |
| CODIGO MUNICIPIO | 50 | 37 | 13 |

**Proyecto de Reforestación
Frecuencias - Empresa de Investigación**

Statistics

| | CODIGO ALDEA | COCINAR EN LEÑA | GAS KEROSEN | GAS PROPANO | CARBON | COCINAR OTROS |
|------------------------|--------------|-----------------|-------------|-------------|--------|---------------|
| N | 50 | 51 | 51 | 51 | 51 | 51 |
| | Valid | 51 | 51 | 51 | 51 | 51 |
| | Missing | 0 | 0 | 0 | 0 | 0 |
| Mean | 3.64 | .94 | .00 | .00 | .00 | 3.92E-02 |
| Mode | 5 | 1 | 0 | 0 | 0 | 0 |
| Std. Deviation | 1.34 | .24 | .00 | .00 | .00 | .20 |
| Variance | 1.79 | 5.65E-02 | .00 | .00 | .00 | 3.84E-02 |
| Skewness | -.099 | -3.865 | .333 | .333 | .333 | 4.893 |
| Std. Error of Skewness | .337 | .333 | 0 | 0 | 0 | .333 |
| Range | 3 | 1 | 0 | 0 | 0 | 1 |
| Sum | 182 | 48 | 0 | 0 | 0 | 2 |

Tabla de Frecuencias

CODIGO ALDEA

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------------|-----------|---------|---------------|--------------------|
| Valid | | | | |
| Pameya | 15 | 29.4 | 30.0 | 30.0 |
| Palama | 11 | 21.6 | 22.0 | 52.0 |
| Saquitacaj | 1 | 2.0 | 2.0 | 54.0 |
| Hacienda María | 23 | 45.1 | 46.0 | 100.0 |
| Total | 50 | 98.0 | 100.0 | |
| Missing | 1 | 2.0 | | |
| Total | 51 | 100.0 | | |

COCINAR EN LEÑA

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 3 | 5.9 | 5.9 | 5.9 |
| Sí | 48 | 94.1 | 94.1 | 100.0 |
| Total | 51 | 100.0 | 100.0 | |

GAS KEROSENE

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 51 | 100.0 | 100.0 | 100.0 |

GAS PROPANO

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 51 | 100.0 | 100.0 | 100.0 |

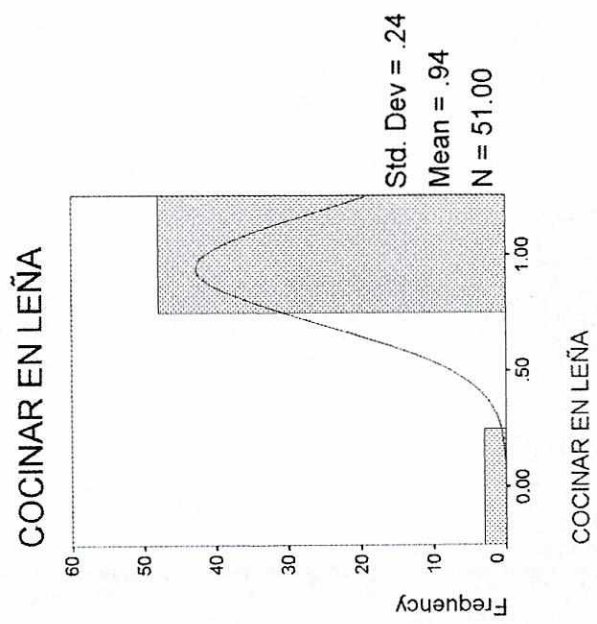
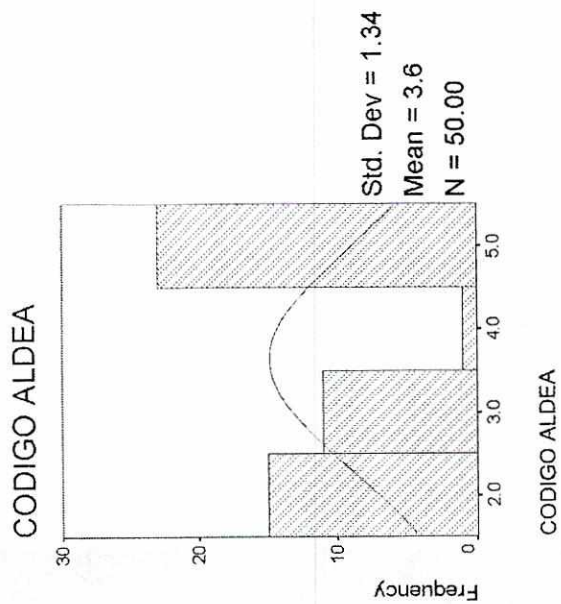
CARBON

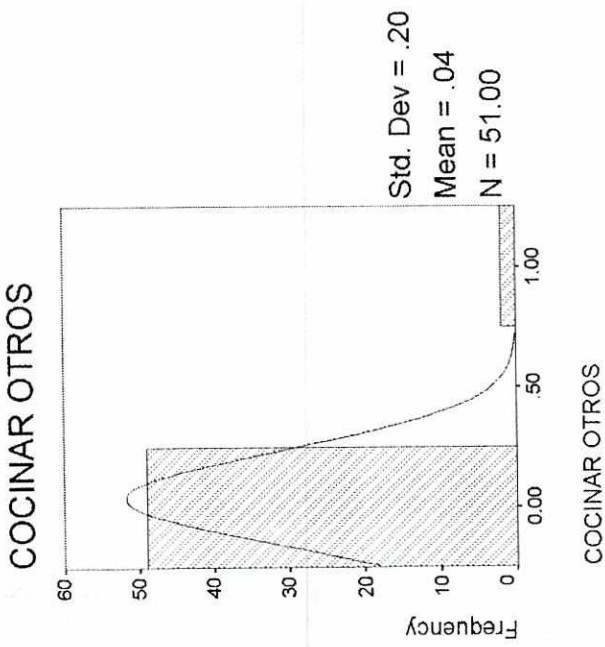
| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 51 | 100.0 | 100.0 | 100.0 |

COCINAR OTROS

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 49 | 96.1 | 96.1 | 96.1 |
| Sí | 2 | 3.9 | 3.9 | 100.0 |
| Total | 51 | 100.0 | 100.0 | |

Histograma





LLUVIA

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 50 | 100.0 | 100.0 | 100.0 |

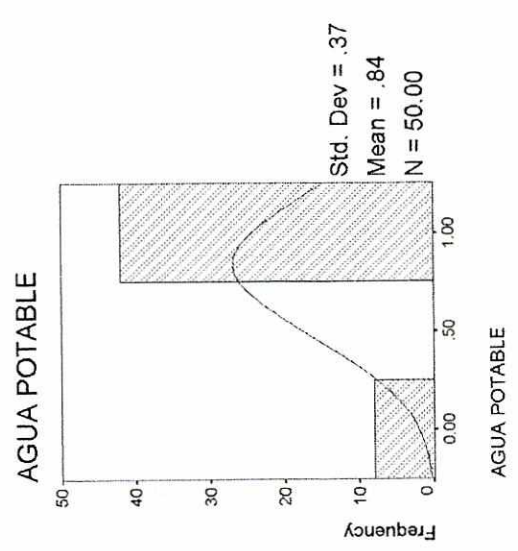
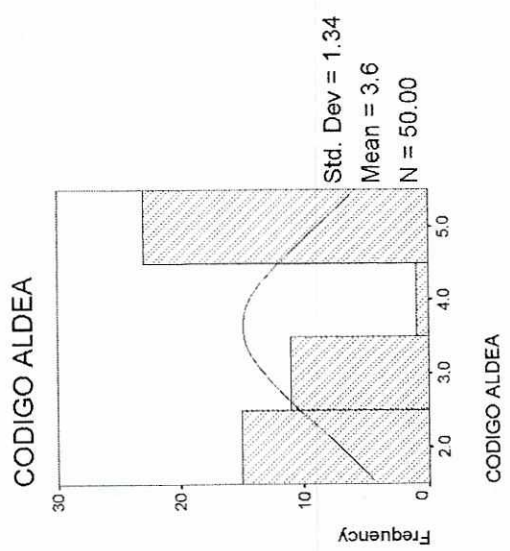
COMPRADA

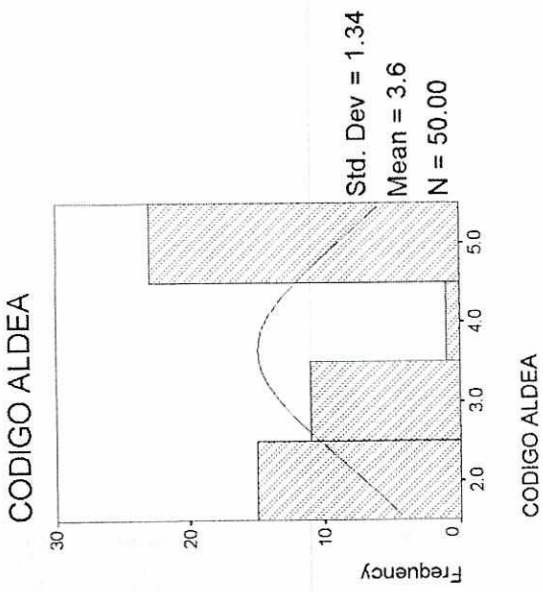
| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 50 | 100.0 | 100.0 | 100.0 |

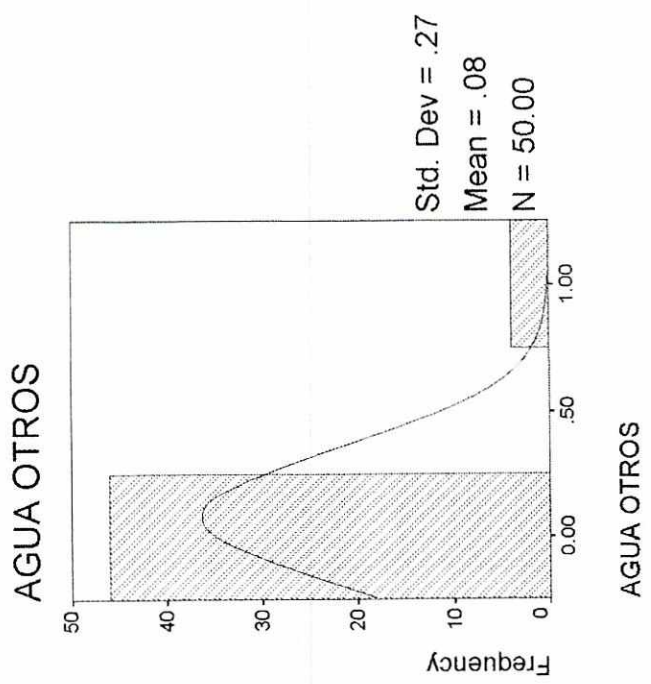
AGUA OTROS

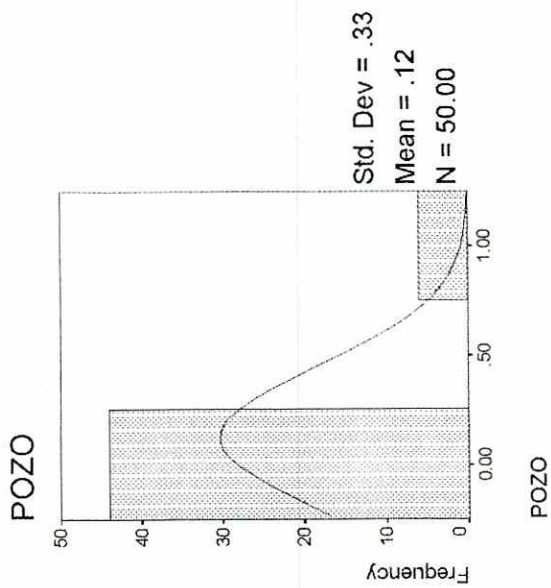
| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------|-----------|---------|---------------|--------------------|
| Valid No | 46 | 92.0 | 92.0 | 92.0 |
| Valid Sí | 4 | 8.0 | 8.0 | 100.0 |
| Total | 50 | 100.0 | 100.0 | |

Histogram









Proyecto de Introducción de Agua Tabla de frecuencias - Empresa de Investigación

Statistics

| | CODIGO ALDEA | AGUA POTABLE | CHORRO PUBLICO | POZO | POZO PUBLICO | RIO | LLUVIA | COMPRADA | OTROS |
|------------------------|--------------|--------------|----------------|-------|--------------|------|--------|----------|-------|
| N | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| Mean | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Median | 3.64 | .84 | .00 | .12 | .00 | .00 | .00 | .00 | .08 |
| Mode | 3.00 | 1.00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| Std. Deviation | 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Variance | 1.34 | .37 | .00 | .33 | .00 | .00 | .00 | .00 | .27 |
| Skewness | 1.79 | .14 | .00 | .11 | .00 | .00 | .00 | .00 | .08 |
| Std. Error of Skewness | -.099 | -1.913 | .337 | 2.412 | .337 | .337 | .337 | .337 | 3.193 |
| Sum | .337 | .337 | .337 | .337 | .337 | .337 | .337 | .337 | .337 |
| | 182 | 42 | 0 | 6 | 0 | 0 | 0 | 0 | 4 |

Tabla de Frecuencia

CODIGO ALDEA

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------------|-----------|---------|---------------|--------------------|
| Valid Pameya | 15 | 30.0 | 30.0 | 30.0 |
| Palama | 11 | 22.0 | 22.0 | 52.0 |
| Saquitacaj | 1 | 2.0 | 2.0 | 54.0 |
| Hacienda Maria | 23 | 46.0 | 46.0 | 100.0 |
| Total | 50 | 100.0 | 100.0 | |

Proyecto de Introducción de Agua Correlaciones bivariadas - Empresa de Investigación

Descriptive Statistics

| | Mean | Std. Deviation | N |
|-----------------|------|----------------|----|
| VIVIENDA PROPIA | .90 | .30 | 50 |
| AGUA POTABLE | .84 | .37 | 50 |

Correlations

| | VIVIENDA PROPIA | AGUA POTABLE |
|-----------------|---------------------|--------------|
| VIVIENDA PROPIA | 1.000 | .218 |
| | Pearson Correlation | .128 |
| | Sig. (2-tailed) | .50 |
| | N | 50 |
| AGUA POTABLE | .218 | 1.000 |
| | Pearson Correlation | .128 |
| | Sig. (2-tailed) | .50 |
| | N | 50 |

**Proyecto de Introducción de Agua
Chi Cuadrada Prueba no paramétrica - Empresa de investigación**

Frecuencias

VIVIENDA PROPIA

| | Observed N | Expected N | Residual |
|-------|------------|------------|----------|
| No | 6 | 25.5 | -19.5 |
| Sí | 45 | 25.5 | 19.5 |
| Total | 51 | | |

AGUA POTABLE

| | Observed N | Expected N | Residual |
|-------|------------|------------|----------|
| No | 9 | 25.5 | -16.5 |
| Sí | 42 | 25.5 | 16.5 |
| Total | 51 | | |

Test Statistics

| | VIVIENDA PROPIA | AGUA POTABLE |
|-------------------------|-----------------|--------------|
| Chi-Square ^a | 29.824 | 21.353 |
| df | 1 | 1 |
| Asymp. Sig. | .000 | .000 |

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25.5.