CAPITULO 1

1. CONCEPTOS ESTADÍSTICOS E INFORMÁTICOS A UTILIZARSE

1.1 HERRAMIENTAS ESTADÍSTICAS

Cuando examinamos un proceso o un fenómeno podemos producir una variada información, entonces es preciso determinar cual es la de interés para los fines que tengamos y como conseguirla, así mismo se debe tener una idea del número de observaciones que son necesarias para disponer de información confiable.

Estadística, es una rama de las matemáticas que se ocupa de reunir, organizar y analizar datos numéricos y que ayuda a resolver problemas como el diseño de experimentos y la toma de decisiones.
Para obtener la información estadística necesaria contamos con dos técnicas estadísticas: Estadística descriptiva e Estadística inferencial.

Antes de explicar estas técnicas, expondremos algunas definiciones necesarias para su comprensión.

Algo importante que hay que mencionar es que no siempre se trabaja con todos los datos. Por ejemplo, resultaría muy costoso obtener los datos de todos los seres humanos, o impráctico (y a la vez destructivo) obtener como datos el tiempo en el que se funden las bombillas producidas por una cierta marca realizando la medición de toda la producción. Por esta razón se considera un subconjunto del total de los casos, sujetos u objetos que se estudian y que se les obtienen los datos.

Definición 1. (Población): Es una colección de elementos acerca de los cuales deseamos hacer alguna inferencia.

Definición 2. (Muestra): Selección de un conjunto de individuos representativos del total de la población objeto de estudio, reunidos como

una representación válida y de interés para la investigación de su comportamiento.

Los criterios que se utilizan para la selección de muestras pretenden garantizar que el conjunto seleccionado represente con la máxima fidelidad a la totalidad de la que se ha extraído, así como hacer posible la medición de su grado de probabilidad.

La muestra tiene que estar protegida contra el riesgo de resultar sesgada, manipulada u orientada durante el proceso de selección, con la finalidad de proporcionar una base válida a la que se pueda aplicar la teoría de la distribución estadística.

Definición 3. (Unidad de muestra o Experimental): Es un elemento de la población.

Tipos de datos

El análisis que se aplique a un conjunto de datos dependerá en gran medida del tipo de datos (Variables) que se quiera analizar. Distinguiremos dos tipos de datos (Variables):

1. **Datos cualitativos**

Son datos cualitativos aquellos que recogen alguna característica no numérica. Ejemplos de variables cualitativas son: el género de un individuo, su provincia o nacionalidad de origen, su estado civil.

2. **Datos cuantitativos**

Son datos que se representan de una forma natural con números. Por ejemplo Altura de una persona, Peso, Ingresos.

1.1.1 Estadística Descriptiva

Una de las ramas de la Estadística más accesible a la mayoría de la población es la Descriptiva. Esta parte se dedica única y exclusivamente al ordenamiento y tratamiento de la información para su presentación por medio de tablas y de representaciones gráficas, así como de la obtención de algunos parámetros útiles para la explicación de la misma. La estadística descriptiva analiza, estudia y describe a la totalidad de individuos de una población. Su finalidad es obtener información, analizarla, elaborarla y simplificarla lo necesario para que pueda ser interpretada cómoda y rápidamente.

Análisis Univariado

Media

La media de un conjunto de n unidades está dada por la suma de todos sus elementos dividido entre el total de elementos.

$$\mu = \frac{\sum x}{N}$$

Donde N es el tamaño de la población y x son los valores de los datos de la población. A la media poblacional se la conoce como μ (mu) y la media muestral como $\overline{x}$ (x barra).

Cálculo de la media muestral $\overline{x}$:

$$\overline{x} = \frac{\sum x}{n}$$

Donde n es el tamaño de la muestra y x son todos los valores que toma la muestra.

Mediana

Medidas de centralización. Colocando todos los valores ordenados en forma creciente o decreciente, la mediana es aquél que ocupa la posición central.

Con un número par de elementos la posición de la mediana esta entre las dos observaciones intermedias y su valor se obtiene promediando dichas observaciones.

Con un numero impar de elementos la posición de la mediana esta dada por $\frac{n+1}{2}$ y el valor de la mediana es el valor que se encuentra en dicha posición.

Moda

Valor que aparece con más frecuencia en un conjunto dado de números. Si son dos los números que se repiten con la misma frecuencia, el conjunto tiene dos modas. Otros conjuntos no tienen moda.

Varianza y Desviación Típica

Estas medidas toman en consideración la forma en que se distribuyen los datos alrededor de la media.

La varianza, S^2, se define como la media de las diferencias cuadráticas de *n* puntuaciones con respecto a su media, es decir

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

La varianza no tiene la misma magnitud que las observaciones (ej. si las observaciones se miden en metros, la varianza lo hace en metros^2). Si queremos que la medida de dispersión sea de la misma dimensionalidad que las observaciones, bastará con tomar su raíz cuadrada positiva. Por ello se define la desviación típica, S, como

$$S = +\sqrt{S^2}$$

Distribución de Frecuencias

Tabla de datos, referentes a una variable en cuestión, en la que se exponen varias categorías de la misma, junto con sus frecuencias o número de veces que se repite en la muestra (puede expresarse también en porcentaje).

Las frecuencias pueden ser absolutas o relativas, simples o acumuladas.

Absoluta:

Es el número de elementos u observaciones de una misma clase.

Relativa:

Es el porcentaje de frecuencia observado en una clase con respecto al total.

Frecuencia Acumulada:

Es la suma de las frecuencias que se encuentran en cada límite. Es decir, el acumulado de la segunda clase es la suma de la primera más la segunda, el acumulado de la tercera clase es la suma de la primera, más la segunda, más la tercera, y así sucesivamente.

La tabla puede tener diferentes formatos y es llamada tabla de frecuencias. Cuando se comparan la frecuencia de dos variables, se compone una tabla de contingencia, en la cual una variable ocupa las filas y la otra las columnas.

Histograma

Se utiliza para representar tablas de frecuencias con datos agrupados en intervalos. Si los intervalos son todos iguales, cada uno de ellos es la base de un rectángulo cuya altura es proporcional a la frecuencia correspondiente.

Si se unen los puntos medios de la base superior de los rectángulos se obtiene el polígono de frecuencias.

En resumen una tabla de frecuencia es una tabla en la que se disponen los datos divididos en grupos ordenados numéricamente, denominados clases

o categorías. El punto medio de la clase se denomina Marca de Clase, la longitud del intervalo de clase es la anchura de clase, una representación grafica de la distribución de frecuencia por medio de barras verticales se denomina histograma de frecuencias, si se unen los puntos medios de los techos del histograma se obtiene el denominado *polígono de frecuencias*.

Diagrama de puntos

Un diagrama de puntos es una forma de resumir datos cuantitativos, en la que cada observación se representa mediante un punto sobre una recta numérica, Si se tuviera muchos datos, cada punto puede representar un número fijo de observaciones.

El diagrama de puntos nos permite apreciar:

a) Localización general de observaciones.
b) Dispersión de las observaciones.
c) Presencia de valores atípicos.

Diagrama de Tallos y Hojas

El diagrama de tallo y hojas se emplea para ilustrar las principales características de los datos y además tiene la ventaja de presentar de manera grafica los valores de los datos.

El diagrama de Tallos y Hojas nos permite apreciar:

a) Localización general de observaciones.

b) Dispersión de las observaciones.

c) Simetría de las observaciones.

Diagrama de cajas

Los diagramas de Cajas y Bigotes (boxplots o box and whiskers), se basan en los siguientes parámetros de la distribución: valor mínimo, los cuartiles: Q 1, Q 2 y Q 3 y el valor máximo.

Los cuartiles son los valores de la variable que dejan por debajo de sí el 25%, 50% y el 75% del total de las puntuaciones y así tenemos por tanto el primer cuartil (Q1), el segundo (Q2)

y el tercer cuartil (Q3).

Por lo tanto tenemos:

$$Q1 = \frac{n+1}{4} \qquad Q2 = \frac{n+1}{2} \qquad Q3 = \frac{3(n+1)}{4}$$

donde *n* es el total del número de datos

El Diagrama de Cajas se construye dibujando un rectángulo desde el primer cuartil hasta el tercer cuartil, la mediana se identifica con una barra vertical dentro de este rectángulo, una línea de extiende desde el tercer

cuartil hasta el valor máximo y otra desde el valor mínimo hasta el primer cuartil.

Ojiva

Una gráfica similar al polígono de frecuencias es la ojiva, pero ésta se obtiene de aplicar parcialmente la misma técnica a una distribución acumulativa y de igual manera que éstas, existen las ojivas mayor que y las ojivas menor que.

Existen dos diferencias fundamentales entre las ojivas y los polígonos de frecuencias (y por esto la aplicación de la técnica es parcial):

1. Un extremo de la ojiva no se "amarra" al eje horizontal, para la ojiva mayor que sucede con el extremo izquierdo; para la ojiva menor que, con el derecho.
2. En el eje horizontal en lugar de colocar las marcas de clase se colocan las fronteras de clase. Para el caso de la ojiva mayor que es la frontera menor; para la ojiva menor que, la mayor.

1.1.2 Estadística Inferencial

Análisis Multivariado

El análisis multivariado incluye los métodos estadísticos que se preocupan por el análisis de las medidas múltiples que se han hecho en un cierto número de individuos.

Estos métodos permiten establecer a partir de datos numerosos, ciertas relaciones y leyes operativas investigar estructuras latentes y ensayar diversas maneras de organizar dichos datos en estructuras y fácilmente utilizables, se logran a partir de dos situaciones completas:

1. Transformándolos y representándolos
2. Reduciéndolos sin perder información

Los valores observados en una unidad experimental no deben influir sobre los valores observados en cualquier otra.

Algunos de los métodos multivariados son: Regresión Múltiple, Análisis Discriminante, Análisis Factorial, entre otros.

Independencia de las Unidades Experimentales

Una condición que deben satisfacer casi todos los métodos multivariados es que las variables medidas en cualquier unidad experimental dada deben ser independientes de las variables semejantes medidas en cualquier unidad experimental, es decir, los valores observados en una unidad experimental no deben influir sobre los valores observados en cualquier otra.

La gran mayoría de los métodos multivariados se originan a partir de la hipótesis de que los datos de variables múltiples provienen de una distribución normal, la mayoría de estos, son robustos con respecto a los datos no normales.

1.1.3 Muestreo

Introducción

En toda investigación estadística existe un conjunto de elementos sobre los que se toma información. Este conjunto de elementos es lo que ese denota con el nombre de población o universo estadístico. Cuando el estadístico o investigador toma información de todos y cada uno de los elementos de la población estadística se dice que esta realizando un censo. Sin embargo esto no es muchas veces posible, ya sea por el costo que resulta de la

toma de información, o bien porque la toma de información lleve consigo la destrucción de los elementos en cuestión, o que la población tenga infinitos elementos, o por otras causas.

Este problema lleva al investigador a tomar la información solo de una parte de los elementos de la población estadística, proceso que recibe el nombre de muestreo.

Existen varios tipos de muestreo, dependiendo de que la muestra sea finita o infinita.

Muestra

Es un subconjunto de tamaño *n* tomada de una población **X** ($x_1, x_2, x_3, \ldots, x_n$).

Muestra aleatoria

Debemos hacer dos consideraciones:

a) **X** es finita;
b) **X** es infinita.

Si **X** es finita, suponer de tamaño *N*; $x_1, x_2, x_3, \ldots, x_n$; tomadas de **X** es aleatoria se cualquier subconjunto de tamaño *n* tomados de **X** tiene igual probabilidad de integrar la muestra.

Si X es infinita; $x_1, x_2, x_3, \ldots, x_n$ es una muestra aleatoria si y solo si $x_1, x_2, x_3, \ldots, x_n$ es un conjunto de variables aleatorias que son independientes idénticamente distribuidas.

Universo

Conjunto de todos los entes cuyas características vamos a investigas.

Universo o Población Investigada

Son todos los entes que están registrados en el instrumento base o marco muestral que utilizaremos para realizar nuestra investigación.

Tipos de muestreo

Muestreo Aleatorio Simple:

La forma más común de obtener una muestra es la selección al azar, es decir, cada uno de los individuos de una población tiene la misma posibilidad de ser elegido. Si no se cumple este requisito, se dice que la muestra es viciada. Para tener la seguridad de que la muestra aleatoria no es viciada, debe emplearse para su constitución una tabla de números aleatorios.

Muestreo Estratificado:

Una muestra es estratificada cuando los elementos de la muestra son proporcionales a su presencia en la población. La presencia de un elemento en un estrato excluye su presencia en otro. Para este tipo de muestreo, se divide a la población en varios grupos o estratos con el fin de dar representatividad a los distintos factores que integran el universo de estudio. Para la selección de los elementos o unidades representantes, se utiliza el método de muestreo aleatorio.

Muestreo por Categorías:

Se divide a la población en estratos o categorías, y se asigna una categoría para las diferentes categorías y, a juicio del investigador, se selecciona las unidades de muestreo. La muestra debe ser proporcional a la población, y en ella deberán tenerse en cuenta las diferentes categorías. El muestreo por categorías se presta a distorsiones, al quedar a criterio del investigador la selección de las categorías.

Muestreo Intencionado:

También recibe el nombre de sesgado. El investigador selecciona los elementos que a su juicio son representativos, lo que exige un conocimiento previo de la población que se investiga.

Muestreo Mixto*:*

Se combinan diversos tipos de muestreo. Por ejemplo: se puede seleccionar las unidades de la muestra en forma aleatoria y después aplicar el muestreo por categorías.

Muestreo Tipo:

La muestra tipo (master simple) es una aplicación combinada y especial de los tipos de muestra existentes. Consiste en seleccionar una muestra "para ser usada" al disponer de tiempo, la muestra se establece empleando procedimientos sofisticados; y una vez establecida, constituirá el módulo general del cual se extraerá la muestra definitiva conforme a la necesidad específica de cada investigación.

Tamaño de la muestra

Calcularemos el tamaño de la muestra teniendo en cuenta que el error de muestreo y el coeficiente de confianza están dados:

$$n = \frac{N\sum_{h=1}^{L} N_h S_h^2}{\frac{e^2}{\lambda_\alpha^2} + \sum_{h=1}^{L} N_h S_h^2},$$

donde λ_α^2 y e^2 son el coeficiente de confianza y el error de muestreo dados respectivamente.

N = Tamaño de la población.

N_h = Tamaño de la *h*-esima muestra

$$S_h^2 = \frac{N_h}{N_h - 1} * P_h Q_h ,$$

donde P_h es el número de individuos que cumplen con la característica de interés y $Q_h = 1 - P_h$.

Calcularemos la constante de proporcionalidad:

$p_h = \frac{N_h}{N}$, que se utilizará como ponderador, teniendo en cuenta que: $\sum p_h = 1$.

Por lo tanto el número de individuos que aleatoriamente se escogerán de la *h*-esima muestra será:

$$n_h = n * p_h$$

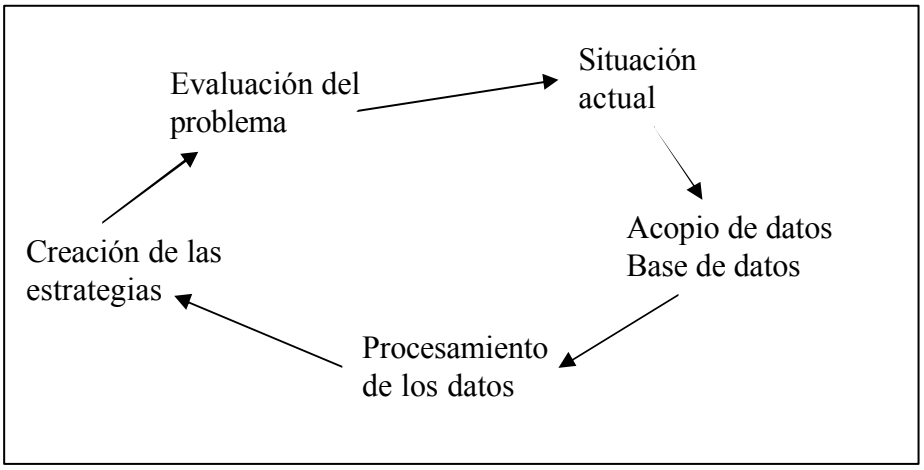
1.2 SISTEMAS DE INFORMACIÓN MÉDICOS

Dada la creciente complejidad y volumen de información que se maneja en el Área de la Salud, situación que desborda cualquier sistema manual, se hace imprescindible contar con sistemas de información que permitan al usuario un mejor control de datos.

1.2.1 Informática Médica

El campo de las ciencias de la Información, que se ocupa del análisis y diseminación de los datos médicos, a través de aplicar la computación a varios aspectos del cuidado de la Salud y la Medicina, es el campo científico que tiene que ver con la información, los datos y el conocimiento biomédico, su almacenamiento, recuperación y uso óptimo para resolver y tomar problemas.

Figura 1.1
Sistemas de Información Médicos



Obstáculos y problemas

a) Se presentan cuando las variables no han sido consideradas desde un principio.
b) El uso de los computadores en la vida laboral no ha estado exento de problemas.
c) La complejidad de la realidad sanitaria, que genera una gran cantidad de datos, obliga a una correcta selección de herramientas informáticas.
d) Para la obtención y manejo de la información es de gran utilidad la computación.
e) Por tanto;

"...no utilizar ordenadores en el área medica, es cerrarse a potenciales mejoras en el manejo de la información generada por los pacientes"

1.2.2 Información médica en Internet

En la era de la información se ha incorporado una nueva palabra que hasta hace poco ni siquiera conocíamos y que ahora es un lugar común, Internet. Aunque en ocasiones no se sepa realmente de qué se trata, cada vez se va haciendo mas evidente, que éste prototipo de la "autopista de la información", es la herramienta que permite acceder y compartir desde

cualquier lugar y a un bajo precio, una cantidad ingente de información hasta ahora ni siquiera vislumbrada.

En estos últimos años, esta red de redes de ordenadores, enlazados por los más diversos medios y utilizando un lenguaje común de comunicaciones, se ha convertido en un nuevo medio global que está siendo aceptado más rápidamente que ningún otro medio de comunicación en la historia.

La rápida expansión, gracias a un incremento de su funcionalidad y facilidad de uso, la ha alejado de sus primeros usuarios exclusivamente militares y científicos, y la ha llevado a incluir una gama de información que oscila desde los temas más desconocidos, hasta los de importancia mundial.

Dentro de esta extensa y heterogénea cantidad de información, la Medicina no podía ser ajena, ya que ésta disciplina es una en las que se usa y necesita mayor cantidad de información, además de la importancia que posee.

Internet y el desarrollo de nuevas tecnologías asociadas, está revolucionando campos de la medicina como la educación médica continuada, investigación, telemedicina, o incluso los sistemas de información y redes internas de los centros sanitarios. Sin embargo,

Internet no es sólo una fuente valiosa para encontrar lo más relevante en investigación y educación médica, sino también, un importante recurso de información en el entorno clínico.

Para muchos médicos, la función más importante de los ordenadores en la búsqueda de información científica, ha sido la exploración en bases de datos bibliográficas de artículos publicados en revistas médicas.

El principal atractivo que ofrece hoy Internet, es el servicio denominado World Wide Web, que proporciona una interfaz gráfica donde la información se presenta en forma de páginas unidas con enlaces hipertexto y por las que es fácil desplazarse para buscar documentos.

Una propuesta, ya real en otras disciplinas, postula la modificación del sistema actual de revisión por expertos o peer review", por un proceso de revisión global de todos los investigadores que consulten los artículos puestos a la consideración pública en Internet. Aunque acogida con ciertas reticencias, puede ser el futuro de la evaluación científica de los artículos médicos.

La base de datos Medline es una de las mejores y más utilizadas del mundo. A través de Intenet se puede acceder a Medline mediante los proveedores tradicionales de servicios on-line o a través de otros suministradores que aprovechan la gran potencialidad de la red.

Lo novedoso es que, gracias a Internet, se puede acceder a Medline de forma gratuita.

Algunos proveedores de información médica en Internet han añadido el uso gratuito de Medline como servicio de valor añadido. En nuestro país empresas privadas o instituciones añaden a su calidad de proveedores de acceso a Internet el uso de Medline sin gasto adicional para sus usuarios.

La cantidad de información presente en Internet es tan inmensa, tan rápidamente creciente y tan poco organizada que se ha dicho, que es poco útil para encontrar información relevante en poco tiempo. Para obviar este problema, en Internet se han creado herramientas que permiten buscar información en la propia red. Índices o directorios que facilitan el acceso a la información organizada de forma jerárquica, siguiendo una determinada clasificación; motores de búsqueda que permiten rastrear grandes volúmenes de información mediante el uso de palabras clave o herramientas que combinan ambas potencialidades o seleccionan los recursos mediante la evaluación de expertos.

Internet no es todavía la solución a los problemas de información planteados por los médicos. La complejidad en las herramientas necesarias para su uso, lo extensivo y complejo de la información contenida en ella, son razones para que el uso de Internet sea aun reducto de entusiastas. A esto se añadiría la incapacidad de los empleadores o

instituciones en las que los médicos trabajan, para proporcionar el equipo necesario en el lugar de trabajo. Sin embargo Internet está cambiando la forma en que las personas acceden a la información y se comunican en todo el mundo.

La forma de acceder a la información médica y los sistemas de información sanitarios se verán modificados a medio plazo, dando una nueva oportunidad para proporcionar un mejor cuidado a nuestros pacientes. La medicina de familia es una especialidad donde la necesidad de acceder a una gran variedad de fuentes de información se une a la dificultad para acceder a colegas o a una buena biblioteca.

1.2.3 Sistemas de Información Médicos en Ecuador

La evolución de la participación del gobierno ecuatoriano en el desarrollo informático, principalmente en la administración pública, desde 1992, presentan algunas actividades que se emprendieron, resultado de un seminario en el que participaron los diferentes sectores protagónicos del quehacer informático en el país, es decir, empresa privada, universidad, sector público.

En la actualidad, como propósito del gobierno de agilitar los procesos y trámites que se deban efectuar en los organismos de la administración pública, se emitió vía decreto ejecutivo, al final de marzo de 1994, la desregulación de funciones del gobierno central, entre las cuales directamente se elimina la necesidad por parte de las entidades de obtener la autorización de la Dirección Nacional de Informática, DINAIN, para proceder a la contratación de adquisición de bienes y/o servicios informáticos, así como también la obligatoriedad de que las empresas proveedoras deban estar previamente registradas.

Como podemos observar en el Ecuador estamos avanzando con la tecnología y alimentándonos con sistemas de información que ayudan al desarrollo del país, uno de estos es el Sistema de Información del Instituto Ecuatoriano de Seguridad Social, con lo cual sus afiliados harán sus trámites fáciles y rápidos.

Figura 1.2
Sistema de Información del Instituto Ecuatoriano de Seguridad Social
IESS

1.3 BASES DE DATOS

Desde tiempos remotos, los datos han sido registrados por el hombre en algún tipo de soporte (piedra, papel, madera, etc.) a fin de que quedara constancia de un fenómeno o idea. Los datos han de ser interpretados para que se conviertan en información útil, esta interpretación supone un fenómeno de agrupación y clasificación.

En la era actual y con el auge de los medios informáticos aparece el almacenamiento en soporte electromagnético, ofreciendo mayores posibilidades de almacenaje, ocupando menos espacio y ahorrando un tiempo considerable en la búsqueda y tratamiento de los datos. Es en este momento donde surge el concepto de bases de datos y con ellas las diferentes metodologías de diseño y tratamiento.

1.3.1 ¿Que es una base de datos?

Una base de datos es una colección de información que existe durante un periodo largo, a menudo de muchos años. En lo colonial, con la expresión BASE DE DATOS se designa una colección que es administrada por un Sistema de administración de base de datos.

El concepto básico en el almacenamiento de datos es el registro. El registro agrupa la información asociada a un elemento de un conjunto, y está compuesto por campos. Así por ejemplo, un registro correspondiente a un libro no es más que un elemento de un conjunto: biblioteca, elenco bibliográfico, etc. A su vez, ese registro contiene toda la información asociada al libro, clasificada en campos: título, autor, fecha de edición, etc.

El objetivo básico de toda base de datos es el almacenamiento de símbolos, números y letras cadentes de un significado en sí, que con un tratamiento adecuado se convierten en información útil. Un ejemplo podría ser el siguiente dato: 19941224, con el tratamiento correcto podría convertirse en la siguiente información: "Fecha de nacimiento: 24 de diciembre de 1994".

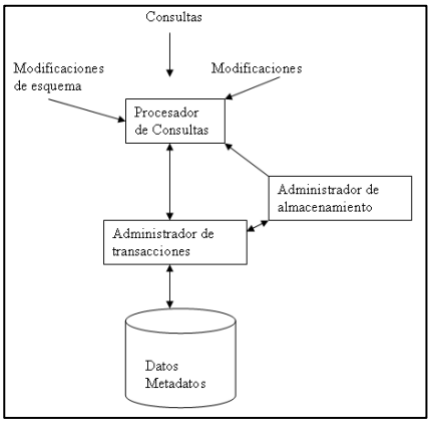
Según van evolucionando los tiempos, las necesidades de almacenamiento de datos van creciendo y con ellas las necesidades de transformar los mismos datos en información de muy diversa naturaleza. Esta información es utilizada diariamente como herramientas de trabajo y

como soporte para la toma de decisiones por un gran colectivo de profesionales que toman dicha información como base de su negocio. Por este motivo el trabajo del diseñador de bases de datos es cada vez más delicado, un error en el diseño o en la interpretación de datos puede dar lugar a información incorrecta y conducir al usuario a la toma de decisiones equivocadas. Se hace necesaria la creación de un sistema que ayude al diseñador a crear estructuras correctas y fiables, minimizando los tiempos de diseño y explotando todos los datos, nace así la metodología de diseño de bases de datos.

Los sistemas de base de datos cambiaron en forma radical después que TED CODD escribió su famoso trabajo en 1970, él propuso que los sistemas deberían presentar al usuario una vista de los datos organizados como tablas llamados relaciones.

1.3.2 Arquitectura de una Base de Datos

Figura 1.3
Arquitectura de una Base de Datos

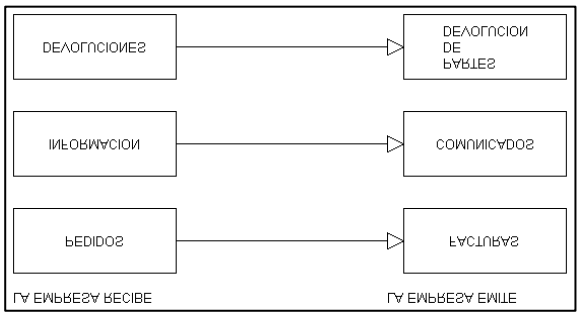


1.3.3 Metodología del Diseño de Datos

La metodología de diseño de datos divide cada modelo en tres esquemas:

A) Modelo Global: Trata de una representación gráfica legible por el usuario y que nos aporta el flujo de información dentro de una organización. No existen reglas para su construcción y se debe realizar siempre el esquema más sencillo posible para la comprensión por parte del usuario de la base de datos. Por ejemplo:

Figura 1.4
Modelo Global



B) Modelo Lógico: Trata de una representación gráfica, mediante símbolos y signos normalizados, de la base de datos. Su objetivo es representar la estructura de los datos y las dependencias de los mismos, garantizando la consistencia y evitando la duplicidad.

C) Modelo Físico: Trata del almacén de los datos, es la base de datos en sí misma, el soporte donde se almacenan los datos y de donde se extraen para convertir los datos en información. En función del gestor de bases de datos empleado, las reglas de almacenamiento varían.

1.3.4 Usuarios de la Base de Datos

En todo sistema de base de datos cabe diferenciar tres tipos de usuarios, entre todos comparten la información pero acceden a ella de una forma diferente, siempre en función de sus necesidades.

El primer grupo de usuarios es el PED (Procesamiento Electrónico de Datos), normalmente compuestos por los operarios de la organización. Las necesidades básicas de este grupo de usuarios son:

a) El foco operativo fundamental se centra en el almacenamiento de los datos, el procesamiento de los mismos y el flujo de datos;
b) Generan informes de tipo listados;
c) Poseen acceso restringido a la información.

El segundo grupo de usuarios es el SIM (Sistemas de Información de Gestión) y suele estar formado por los mandos medios de la organización. Las necesidades básicas de este grupo de usuarios son:

a) El foco operativo se fundamenta en la toma de decisiones, tomando como partida los datos del grupo PED e introduciendo un volumen pequeño de información;
b) No poseen acceso medianamente restringido a la información;
c) Generan informes de resúmenes de datos del grupo PED y listados de la información que introducen.

El tercer y último grupo de usuarios lo forman el STD (Sistema de apoyo a Toma de Decisiones), este grupo se centra en el nivel más alto de la organización y poseen las características siguientes:

a) El foco operativo se centra en la decisión, con una entrada mínima de datos;
b) No tienen acceso restringido;
c) Generan informes globales que les sirven como apoyo a las tomas de decisiones del negocio, estos son los informes más importantes y suelen ir acompañados de resúmenes, gráficas y sobre todo centrados en la evolución y comparación de la información.

Cabe destacar la figura de un cuarto grupo de usuarios, en este caso usuarios avanzados, que está compuesto por los administradores del sistema, cuya opinión es fundamental para seleccionar el soporte de los datos, evitar la duplicación de información ya existente en otros sistemas

y sobre todo puede aportar el conocimiento de sus usuarios, sus necesidades y los problemas ya resueltos.

1.3.5 Ciclo de Vida de una Base de Datos

El ciclo de vida de un desarrollo de una base de datos consta de siete pasos:

1. Análisis de las necesidades;
2. Estudio de viabilidad;
3. Definición de requisitos;
4. Diseño conceptual / lógico;
5. Implementación;
6. Evaluación y Mantenimiento

Análisis de las necesidades

En reunión con el cliente se deben documentar los tres grupos de usuarios, las necesidades de información de cada uno de ellos, así como los informes que cada uno necesita para su actividad y el contenido de los mismos. Cuanta más precisión exista en estos requisitos iniciales más preciso será el desarrollo de la base de datos, debe quedar documentados los niveles de seguridad de los grupos de usuarios, los derechos de cada uno de ellos sobre los datos, los requisitos de los sistemas informáticos del cliente (sistema operativo, tipo

de red, servidores, etc.) y la ubicación de los usuarios. No hay que olvidar que en ocasiones existen ya sistemas de almacenamiento de datos, por tanto es conveniente analizar los datos ya existentes y analizar las posibles relaciones con la base de datos a desarrollar.

Estudio de viabilidad

Un estudio de viabilidad implica la preparación de un informe con las características siguientes:

1. Viabilidad tecnológica. ¿Hay tecnología suficiente para el desarrollo?.

2. Viabilidad operacional. ¿Existen suficientes recursos humanos, presupuesto, experiencia y formación para el desarrollo?.

3. Viabilidad económica. ¿Se pueden identificar los beneficios? ¿Los beneficios costearían el desarrollo del sistema? ¿Se pueden medir los costos y los beneficios?.

Definición de requisitos

Los requisitos de desarrollo involucran el software y hardware necesario para la implementación, los recursos humanos necesarios (tanto internos como externos), la formación al personal.

Conviene señalar quienes van a ser los interlocutores y fijar un calendario de reuniones, actividades y del proyecto.

Hay que definir la figura de la persona que será la encargada de velar en cada momento que no se está rebasando el alcance del proyecto, así como asegurar que la implementación está encaminada a subsanar las necesidades del cliente.

Diseño

En esta etapa se crea un esquema conceptual de la base de datos. Se desarrollan las especificaciones hasta el punto en que puede comenzar la implementación. Durante esta etapa se crean modelos detallados de las vistas de usuario y sobre todo las relaciones entre cada elemento del sistema, documentando los derechos de uso y manipulación de los diferentes grupos de usuarios.

Si parte de la información necesaria para crear algún elemento establecido ya se encuentra implementado en otro sistema de

almacenamiento hay que documentar que relación existirá entre uno y otro y detallar los sistemas que eviten la duplicidad o incoherencia de los datos.

El diseño consta, como se vio anteriormente, de tres fases: el diseño global o conceptual, el diseño lógico y el modelo físico.

Implementación

Una vez detallado el modelo conceptual se comienza con la implementación física del modelo de datos, a medida que se va avanzando en el modelo el administrador del sistema va asegurando la corrección del modelo.

La implementación consiste en el desarrollo de las tablas, los índices de los mismos, las condiciones de validación de los datos, la relación entre las diferentes tablas. Por otro lado, la definición de las consultas y los parámetros a utilizar por cada una de ellas.

Una vez finalizada la implementación física, se asignan las correspondientes medidas de seguridad y se ubica la base de datos en el lugar correspondiente.

Evaluación y Perfeccionamiento

En esta última etapa todos los usuarios del sistema acceden a la base de datos y deben asegurarse el correcto funcionamiento de la misma, que sus derechos son los adecuados, teniendo a su disposición cuanta información necesiten. También deberán asegurarse que el acceso a los datos es cómodo, práctico, seguro y que se han eliminado, en la medida de lo posible, las posibilidades de error.

El administrador se asegura que todos los derechos y todas las restricciones han sido implementadas correctamente y que se ha seguido en manual de estilo en la totalidad de la implementación.

1.3.7 El Lenguaje de Gestión de Base de datos

Hasta la década de los 80, las personas que preparaban las consultas e informes de una base de datos debían ser programadores. Al aparecer las bases de datos con lenguajes de consulta sencillos y estandarizados, semejantes al lenguaje natural, el proceso de consulta puede hacerlo cualquier usuario mediante un lenguaje escrito asequible.

El lenguaje de gestión de bases de datos más conocido en la actualidad es el ***SQL, Structured Query Language,*** que es un lenguaje estándar internacional, comúnmente aceptado por los fabricantes de generadores de bases de datos.

El lenguaje de consulta estructurado (SQL) es un lenguaje de base de datos normalizado, utilizado por los diferentes motores de bases de datos para realizar determinadas operaciones sobre los datos o sobre la estructura de los mismos. Pero como sucede con cualquier sistema de normalización hay excepciones para casi todo; de hecho, cada motor de bases de datos tiene sus peculiaridades y lo hace diferente de otro motor, por lo tanto, el lenguaje SQL normalizado no nos servirá para resolver todos los problemas, aunque si se puede asegurar que cualquier sentencia escrita en ANSI será interpretable por cualquier motor de datos.

El SQL trabaja con estructura cliente/servidor sobre una red de ordenadores. El ordenador cliente es el que inicia la consulta; el ordenador servidor es que atiende esa consulta. El cliente utiliza toda su capacidad de proceso para trabajar; se limita a solicitar datos al ordenador servidor, sin depender para nada más del exterior. Estas peticiones y las respuestas son transferencias de textos que cada ordenador cliente se encarga de sacar por pantalla, presentar en informes tabulados, imprimir, guardar, etc., dejando el servidor libre.

El SQL permite:

a) Definir una base de datos mediante tablas.
b) Almacenar información en tablas.
c) Seleccionar la información que sea necesaria de la base de datos.

d) Realizar cambios en la información y estructura de los datos.

e) Combinar y calcular datos para conseguir la información necesaria.

SQL es el lenguaje de comunicación entre el programa cliente y programa servidor, el motor de base de datos es un programa servidor, en el que está la base de datos propiamente dicha.

El usuario accede con alguno de los programas cliente disponibles para realizar requerimientos al motor.

SQL se puede emplear para:

a) Crear y modificar la estructura de una tabla de datos.

b) Seleccionar información de una tabla.

c) Añadir datos a una tabla.

d) Introducir información en una tabla.

e) Realizar consultas entre tablas con campos comunes.

Existen tres tipos de comandos SQL:

a) DDL **Data Definition Language,** que permiten crear y definir nuevas bases de datos, campos e índices.

b) DML **Data Manipulation Language,** que permiten generar consultas para ordenar, filtrar y extraer datos de la base de datos.

c) DCL **Data Control Language,** contiene elementos útiles para trabajar en un entorno multiusuario en el que es importante la

protección de la información, la seguridad de las tablas y el establecimiento de restricciones de acceso.

Comandos DDL	
Comando	Descripción
CREATE	Utilizado para crear nuevas tablas, campos e índices
DROP	Empleado para eliminar tablas e índices
ALTER	Utilizado para modificar las tablas agregando campos o cambiando la definición de los campos.
Comandos DML	
Comando	Descripción
SELECT	Utilizado para consultar registros de la base de datos que satisfagan un criterio determinado
INSERT	Utilizado para cargar lotes de datos en la base de datos en una única operación.
UPDATE	Utilizado para modificar los valores de los campos y registros especificados
DELETE	Utilizado para eliminar registros de una tabla de una base de datos
Comandos DCL	
Comando	Descripción
GRANT	Concede privilegios o papeles, roles, a un usuario o a otro rol.
TRUNCATE	Elimina todas las filas de una tabla.
REVOKE	Retira los privilegios de un usuario o rol de la base de datos.

SENTENCIAS DML

Las sentencias DML son aquellas sentencias que me van a permitir manipular los datos, van a modificar el contenido de los datos de la base de datos.

Los cambios que provocan estas sentencias no ocurren al momento, sino que tienen lugar cuando nosotros aceptamos la transacción. Esta aceptación de la transacción se realiza mediante la orden COMMIT. El funcionamiento es el siguiente: Cuando se modifican los datos, se

guarda esta modificación en la estructura interna de la base de datos, en los segmentos que se denominan de Rollback, en ese mismo momento se produce un bloqueo en el registro que se está modificando para evitar que otros usuarios puedan modificarlo. Cuando se produce el commit, los datos se almacenan en la base de datos y permite a otros usuarios acceder a los mismos. Si por otra parte, decidimos deshacer los cambios que se han producido, antes de realizar un COMMIT, tendremos que hacer un ROLLBACK de los datos que han sido modificados, haciendo esto conseguimos que se recuperen los mismos datos que había antes de la modificación.

La sentencia INSERT

La sentencia insert lo que hará será añadir un nuevo registro dentro de la base de datos, su sentencia básica es la siguiente:

```
INSERT INTO tabla (campos) VALUES (valores);
```

Siendo el parámetro campos opcional, que lo utilizaremos cuando:

1- No vayamos a insertar tantos valores como campos tiene la tabla, con lo cual le tendremos que indicar que campos quiero rellenar

2- No vayamos a insertar los valores en el orden que tienen los campos en la tabla.

La sentencia UPDATE

Esta sentencia me va a permitir actualizar valores dentro de los diferentes registros de una tabla, su sintaxis básica es:

```
UPDATE tabla SET campo=valor WHERE condición
```

La sentencia DELETE

Me va a permitir eliminar registros de una tabla. Su sentencia es:

```
DELETE FROM tabla WHERE condición
```

Tanto en las sentencias delete, como en las de update es importante que tengamos en cuenta de especificar la cláusula WHERE, porque sino lo que tendremos es que se van a actualizar todos los registros de la tabla.

JOINS

Los joins son la manera que tendremos nosotros para obtener datos de diferentes tablas. Los joins me van a permitir relacionar las claves

primarias con sus claves foráneas correspondientes. Tenemos diferentes tipos de joins:

INNER JOIN: Agarra de la primera tabla todos los registros que estén relacionados con registros de la tabla a la que se va a unir (segunda tabla), y de la tabla segunda todos aquellos registros que estén relacionados con registros de la primera tabla. Es decir, valores que no estén relacionados no me los va a mostrar.

LEFT JOIN: de la primera tabla cogerá todos los registros estén o no relacionados con registros de la tabla de la derecha, mientras que de la tabla de la derecha sólo cogerá aquellos registros que estén relacionados con registros de la tabla de la izquierda

RIGHT JOIN: es el caso contrario del left join.

FULL JOIN: cogerá todos los registros de la tabla de la izquierda, estén o no relacionados con los de la tabla de la derecha, mientras que de la tabla de la derecha cogerá todos los registros estén o no relacionados con registros de la tabla de la izquierda.

CROSS JOIN: hará un producto cartesiano cogiendo los registros de la tabla izquierda y relacionándolos individualmente con los registros de la tabla derecha también de forma individual, y hará lo mismo con los registros de la tabla de la derecha.

La sentencia para estos joins sería como sigue:

```
SELECT T1.C1, T1.C2, T2.C2 FROM T1, T2
WHERE T1.PK=T2.FK --> INNER JOIN

SELECT T1.C1, T1.C2, T2.C2 FROM T1, T2
WHERE T1.PK=T2.FK (+)--> RIGHT JOIN

SELECT T1.C1, T1.C2, T2.C2 FROM T1, T2
WHERE T1.PK(+)=T2.FK --> LEFT JOIN
```

Procedimiento Almacenado

Un procedimiento almacenado (stored procedure) no es más que una colección de sentencias Transact SQL (el dialecto SQL de SQL Server 6.5) que se constituye como si de una función de un lenguaje estructurado (C, Pascal) se tratase. Es decir, es posible llamarlo mediante un identificador, puede recibir argumentos y devolver un valor de retorno.

Transact SQL, es una versión extendida del lenguaje ANSI SQL que posee características que hacen que puedan definirse pseudo funciones. Así, y a modo de ejemplo, es posible definir variables, posee estructuras de control de flujo, etc.

La característica primordial de los procedimientos almacenados es que se optimizan en el momento de su creación. Esto supone que, a diferencia de

lo que sucede con las sentencias SQL que se envían el gestor de manera interactiva, los procedimientos almacenados pasan previamente por un proceso de normalización.

Cuando se crea un procedimiento almacenado el procesador de consultas del gestor crea una versión del mismo con una cierta estructura normalizada, y la almacena en una de las tablas de sistema. Las siguientes ejecuciones de dicho procedimiento, no necesitarán consumir el tiempo necesario para llevar a cabo este proceso de normalización, con lo que su ejecución será más rápida.

Por otra parte, cuando el procedimiento se ejecuta por vez primera, se produce su compilación y la optimización del acceso del procedimiento a los datos. Este proceso optimizado se mantiene en memoria para posteriores ejecuciones con el consiguiente ahorro adicional de tiempo y recursos.

Sentencias SQL del procedimiento almacenado

El cuerpo del procedimiento estará integrado por un conjunto de sentencias SQL que realizarán las tareas que esperamos del mismo y que se especificarán en la definición del procedimiento siguiendo a la cláusula AS

En general podemos decir que en un procedimiento almacenado pueden incluirse cualquier número y tipo de sentencias Transact SQL. Sin

embargo, es necesario comentar algunas restricciones respecto a la creación de objetos No pueden incluirse las siguientes sentencias CREATE

1. CREATE VIEW
2. CREATE TRIGGER
3. CREATE DEFAULT
4. CREATE PROCEDURE
5. CREATE RULE

1.4 LENGUAJES DE PROGRAMACIÓN

Los lenguajes de programación son herramientas que nos permiten crear programas y software. Entre ellos tenemos Delphi, Visual Basic, Pascal, Java, DreamWeaver, FrontPage etc.

Una computadora funciona bajo control de un programa el cual debe estar almacenado en la unidad de memoria; tales como el disco duro.

Los lenguajes de programación de una computadora en particular se conoce como código de máquinas o lenguaje de máquinas, estos lenguajes codificados en una computadora específica no podrán ser ejecutados en otra computadora diferente.

Para que estos programas funcionen para diferentes computadoras hay que realizar una versión para cada una de ellas, lo que implica el aumento del costo de desarrollo.

Por otra parte, los lenguajes de programación en código de máquina son verdaderamente difíciles de entender para una persona, ya que están compuestos de códigos numéricos sin sentido nemotécnico.

Los lenguajes de programación facilitan la tarea de programación, ya que disponen de formas adecuadas que permiten ser leídas y escritas por personas, a su vez resultan independientes del modelo de computador a utilizar; representan en forma simbólica y en manera de un texto los códigos que podrán ser leídos por una persona; son independientes de las computadoras a utilizar.

Existen estrategias que permiten ejecutar en una computadora un programa realizado en un lenguaje de programación simbólico. Los procesadores del lenguaje son los programas que permiten el tratamiento de la información en forma de texto, representada en los lenguajes de programación simbólicos.

1.4.1 Programación Web

La programación Web, parte de las siglas WWW, que significan World Wide Web o telaraña mundial.

Para realizar una página con la programación Web, se deben tener claros, tres conceptos fundamentales los cuales son, el URL (Uniform Resource Locators), es un sistema con el cual se localiza un recurso dentro de la red, este recurso puede ser una pagina web, un servicio o cualquier otra cosa. En resumen el URL no es más que un nombre, que identifica una computadora, dentro de esa computadora un archivo que indica el camino al recurso que se solicita.

El siguiente concepto dentro de la programación Web, es el protocolo encargado de llevar la información que contiene una pagina Web por toda la red de Internet, como es el HTTP (Hypertext Transfer Protocol).

Y por ultimo el lenguaje necesario cuya funcionalidad es la de representar cualquier clase de información que se encuentre almacenada en una pagina Web, este lenguaje es el HTML (Hypertext Markup Language).

En la programación Web, el HTML es el lenguaje que permite codificar o preparar documentos de hipertexto, que viene a ser el lenguaje común para la construcción de una pagina Web.

Con el comienzo de Internet y la programación Web, se desfasaron los diseños gráficos tradicionales, con lo que se empezaron a diseñar interfaces concretas para este medio, buscando ficheros pequeños para facilitar la carga de los mismos. La programación web se orientaba a un diseño muy cargado e interactuando con el usuario, mientras que al empezar a competir con millones de webs se ha optado más por el diseño sencillo y de fácil comprensión.

En programación web se creó la necesidad de conocer a fondo diferentes lenguajes de programación como HTML, JavaScript y DHTML. Con esto se creó un nuevo profesional de la informática, el diseñador web, experto en estos menesteres, que viene siendo algo así como un experto en programación web, dando una diferencia entre el diseñador gráfico tradicional y el programador de aplicaciones llevadas a Internet.

Los lenguajes de programación que utilizaremos en el desarrollo del sistema son:

DREAMWEAVER

DreamWeaver es un editor de código HTML, basado en un editor visual que nos permite diseñar páginas web de un modo muy sencillo sin saber programar directamente en código.

Dreamweaver MX 2004 es la opción profesional para la creación de sitios y aplicaciones web. Proporciona una combinación potente de herramientas visuales de disposición, características de desarrollo de aplicaciones y soporte para la edición de código. Gracias a las robustas características para la integración y diseño basado en CSS, Dreamweaver permite que los diseñadores y desarrolladores web creen y manejen cualquier sitio web con toda facilidad.

MACROMEDIA FLASH MX

Flash MX es un programa diseñado para crear elementos y películas multimedia interactivas muy comprimidas, que podemos ver en todo tipo de páginas y sitios Web de Internet.

Macromedia Flash MX permite a los diseñadores y desarrolladores integrar video, texto, audio y gráficos en experiencias dinámicas que le permiten al cliente imbuirse en su vivencia y que producen resultados superiores para marketing y presentaciones interactivas, aprendizaje electrónico e interfaces de usuario de aplicaciones. Flash es la plataforma de software de predominancia indiscutible, usada por más de un millón de profesionales y con una presencia en más del 97% de los equipos de escritorio con conexión a Internet en todo el mundo, así como en una amplia gama de dispositivos.

MACROMEDIA FIREWORKS 3.0

Fireworks MX 2004 tiene las herramientas que los profesionales del web necesitan para crear de todo, desde botones gráficos sencillos hasta sofisticados efectos de rollover. Importe, edite e integre fácilmente todos los principales formatos gráficos, incluidos imágenes vectoriales y de mapas de bits.

Un "rollover", es básicamente una técnica por la cual una imagen es reemplazada por otra en función a un evento del usuario.

Active Server Pages (Programación dinámica de páginas web)

ASP es una tecnología creada por Microsoft mediante la cual se pueden construir páginas webs dinámicas.

Con el HTML podemos programar fantásticas webs, si además utilizamos ASP podremos dotar a nuestro Web site de todo el dinamismo que necesitemos. Podremos procesar datos del usuario en tiempo real tales como formularios, consultas a bases de datos, etc.

1.4.2. Costos de Licencias

	Microsoft SQL Server 2000 Enterprise Edition License Only Microsoft - 81000628	desde **$697.00**
	Microsoft SQL Server 2000 Standard Edition 5 CAL Microsoft - 22800753	desde **$858.00**
	MacroMedia Dreamweaver MX 2004 Commercial Completo, PC, Mac MacroMedia - Misc. Programación - DWD070D000	desde **$295.00**
	MacroMedia Fireworks MX 2004 Commercial Full Product, PC/Mac MacroMedia - Misc. Programación - FWD070D000	desde **$240.00**
	MacroMedia UPG Flash MX Professional 2004 Upgrade MacroMedia - PFD070D100	desde **$253.26**