

98  
2y



**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

**FACULTAD DE INGENIERIA**

**SISTEMA DE APOYO A TOMA DE DECISIONES  
EN INFERENCIA ESTADISTICA**

**TESIS PROFESIONAL**

QUE PARA OBTENER EL TITULO DE  
**INGENIERO EN COMPUTACION**

P R E S E N T A  
**JOSE LUIS PICHARDO LOPEZ**

DIRECTOR  
**M. I. VICTOR FLORES ZAVALA**

**TESIS CON  
FALLA DE ORIGEN**



**MEXICO, D. F.**

**1996**

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México



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

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

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

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

***Sistema de Apoyo a Toma de Decisiones en Inferencia Estadística***

***Tesis Profesional  
Ingeniería en Computación  
José Luis Pichardo López***

***Director  
M.I. Víctor Flores Zavala***

***Facultad de Ingeniería  
Universidad Nacional Autónoma de México***

A mi Padre y a mi Madre

A mi Hermana y a mis Hermanos

A Lilia

*"Lo importante no es lo que hicieron de  
nosotros, sino lo que nosotros hacemos  
con eso que hicieron de nosotros"*

Jean-Paul Sartre

# Contenido

## **1 Introducción**

<b>VISIÓN GENERAL</b>	<b>1</b>
<b>TEORÍA DE SISTEMAS</b>	<b>2</b>
Elementos de un sistema	
Propósito común para alcanzar un objetivo	
Análisis y diseño de sistemas	
<b>IMPACTO DE LAS COMPUTADORAS</b>	<b>5</b>
Impacto en las organizaciones	
Impacto en los individuos	
<b>EL SOFTWARE</b>	<b>8</b>
Características del Software	
Proceso del Software o Componentes del Software	
Ingeniería de sistemas de computadora	
Sistemas basados en computadora	
El elemento software	
Aplicaciones del Software	
Importancia del Software	

## **2 Sistemas de información**

<b>INTRODUCCIÓN</b>	<b>16</b>
Sistemas de información	
<b>EVOLUCIÓN DE LOS SISTEMAS DE INFORMACIÓN</b>	<b>18</b>
Los datos, foco inicial de atención	
El nuevo foco de atención de la información	
El interés por el soporte a decisiones	
<b>SISTEMAS DE APOYO A DECISIONES</b>	<b>20</b>
Definición de un DSS	
¿Qué es un DSS?	
Objetivo de un DSS	
Características de un DSS	
Origen de los DSS	
Tipos de decisiones según Simon	
Fases de la toma de decisiones de Simon	
Niveles de toma de decisiones	
Tipos de sistemas para decisiones	
Estructura del problema	
¿Qué son los Sistemas Expertos?	
Características de un sistema experto	
Beneficios de los sistemas expertos	
Desventajas	
Relaciones con otro software	
Tendencias	
Software de apoyo a toma de decisiones	
<b>SISTEMAS DE INFORMACIÓN COMO ESTRATEGIA</b>	<b>32</b>
¿Qué significa estratégico?	

Rol de los sistemas de información en las estrategia  
 ¿Qué son los sistemas estratégicos?  
 Manejo estratégico  
 Planeación de sistemas de información  
 Costos y beneficios de la planeación  
 El proceso de planeación  
 Metas y objetivos  
 Tecnología de información como agente de cambio organizacional

### ***3 Estadística***

¿QUÉ ES Y PARA QUÉ SIRVE LA INFERENCIA ESTADÍSTICA?	39
MODELOS PROBABILÍSTICOS	40
DISTRIBUCIONES DISCRETAS	40
Proceso o secuencia de Bernoulli	
Distribución Bernoulli	
Distribución binomial	
Distribución Pascal	
Distribución de Poisson	
DISTRIBUCIONES CONTINUAS	43
Distribución uniforme o rectangular	
Distribuciones normal y normal estándar	
Distribución exponencial	
INFERENCIA ESTADÍSTICA	45
Estimación de parámetros	
Estimadores puntuales	
Propiedades deseables de los estimadores puntuales	
Estimadores insesgados	
Estimador eficiente	
Estimador consistente	
Intervalos de confianza	
PRUEBA DE HIPÓTESIS ESTADÍSTICA	49
Hipótesis estadística	
Conceptos básicos para la prueba de hipótesis	
Tipos de errores y sus probabilidades	
Procedimiento de la prueba de hipótesis	
Dependencia entre $\alpha$ , $\beta$ y el tamaño de la muestra	
INFERENCIA ESTADÍSTICA PARA MEDIAS	54
Tipos de pruebas para medias	
Prueba de medias para una muestra con desviación estándar conocida	
Prueba de medias para una muestra con desviación estándar desconocida y muestra pequeña	
Prueba de medias para dos muestras independientes con desviación estándar conocida	
Prueba de medias para dos muestras independientes con desviación estándar desconocida y muestras pequeñas	
Prueba de medias para muestras pareadas	
INFERENCIA ESTADÍSTICA PARA VARIANCIAS	59
Tipos de prueba para variancias	
Prueba de variancias para una muestra	
Prueba de variancias para dos muestras	

Relación entre las pruebas de medias y la de variancias	
<b>INFERENCIA ESTADÍSTICA PARA PROPORCIONES</b>	<b>60</b>
Tipos de pruebas para proporciones	
Prueba de proporciones para una muestra	
Prueba de proporciones para dos muestras	
<b>INFERENCIA ESTADÍSTICA PARA BONDAD DE AJUSTE</b>	<b>62</b>
Tipos de pruebas para bondad de ajuste	
$\chi^2$ Prueba de bondad de ajuste para distribuciones discretas	
Corrección de mínima frecuencia para el estadístico $\chi^2$	
$\chi^2$ Prueba de bondad de ajuste para distribuciones continuas	

#### **4 Análisis y desarrollo del sistema**

<b>CICLO DE VIDA DE LOS SISTEMAS</b>	<b>64</b>
<b>ANÁLISIS</b>	<b>65</b>
Identificación de las necesidades	
Estudio de factibilidad	
Análisis de requerimientos	
Tareas de análisis	
<b>DISEÑO</b>	<b>68</b>
Diseño preliminar	
Diseño detallado	
Diseño y calidad del Software	
<b>CODIFICACIÓN Y LENGUAJE DE PROGRAMACIÓN</b>	<b>88</b>
Microsoft Excel	
Macros	
Codificación	
<b>PRUEBAS</b>	<b>100</b>
Objetivo de la prueba	
Información FI UNAM	
Parámetros Estadística Descriptiva	
Modelos Probabilísticos	
Pruebas de Bondad de Ajuste	
Pruebas de Hipótesis Medias	
Pruebas de Hipótesis Variancias	
Pruebas de Hipótesis Proporciones	
<b>CONCLUSIONES Y RECOMENDACIONES</b>	<b>141</b>
<b>BIBLIOGRAFÍA</b>	



# 1

---

## *Introducción*

### **VISIÓN GENERAL.**

□ Durante las tres décadas pasadas, una nueva y dinámica actividad ha venido siendo desarrollada en las organizaciones: el diseño, operación, y mantenimiento de sistemas de información basados en computadora.

Mientras muchas actividades y recursos organizacionales han sido dedicados, por largo tiempo, a la adquisición y al proceso de información, la llegada de las computadoras electrónicas ha logrado incrementar en mucho las capacidades de este proceso. Los sistemas de información basados en computadora han influido en organizaciones de todos tipos y tamaños. De este modo, muchas organizaciones poseen o rentan computadoras; otras obtienen beneficios de otros negocios dedicados a la prestación de estos servicios.

Pero, ¿Qué es un sistema de información? Para nuestros propósitos, se define un sistema de información como un conjunto de procedimientos organizados que, cuando son ejecutados, proveen de información que apoya el control de la organización y/o la toma de decisiones de la misma.

La información es una entidad tangible, o bien intangible, que reduce la incertidumbre acerca de algún estado o evento. Los sistemas de información nos rodean por completo. Las organizaciones utilizan los sistemas de información para procesar transacciones, reducir costos y, finalmente, para generar ganancias de sus productos o servicios.

Los sistemas de información basados en computadora son complejos y a menudo vistos como rígidos y difíciles de cambiar. La presencia de tecnología de computadoras es ocasionalmente intimidante para los usuarios. Los usuarios son los miembros de la organización que utilizan los sistemas de información.

Por otro lado, el departamento de sistemas tiene por actividades: el diseño, desarrollo y mantenimiento de los sistemas de información. El diseñador de sistemas trabaja con los usuarios para crear procedimientos, algoritmos, estructuras, así como archivos y su contenido. El analista y el programador se encargan de la implantación y la codificación del nuevo sistema, así como de preparar la documentación que describe a este, y después, mantenerlo operativo. Pero, ¿Cuál es la misión de un sistema de información? Bien, en los primeros días de el proceso de información, los sistemas de información literalmente trabajaban como una "fábrica de papel", para obtener la nomina de empleados, las facturas de los clientes y el inventario de productos. Durante esa era, los objetivos estaban definidos por la medición de la productividad de la computadora, el tiempo de servicio ininterrumpido de la máquina, el número de transacciones procesadas en cierto lapso de tiempo, la cantidad de líneas impresas, etc.

Después, en la era de el manejo de sistemas de información, el centro de atención era la generación de reportes para los niveles gerenciales o los reportes sumariados para los niveles más altos.

La era actual ha regresado a un objetivo clásico dentro de cualquier organización: Proporcionar información correcta a la persona correcta y en el momento correcto. Mejorar la calidad y oportunidad del desempeño de todos los participantes de la organización, mediante la utilización de los sistemas de información. El objetivo es mejorar el desempeño. El foco de atención es la gente que conforma la organización, y la fuente con la cual se debe de nutrir la misión, es la tecnología de información.

## TEORÍA DE SISTEMAS

El campo del análisis y diseño de sistemas de información tiene sus bases en la teoría general de sistemas. La teoría general de sistemas enfatiza la necesidad de examinar todas las partes de un sistema. Además de enfocar su atención a todas las partes, la teoría general de sistemas ayuda a la comunicación entre especialistas de diferentes campos de acción.

Un sistema de información involucra personas que pertenecen a diferentes niveles dentro de la organización. Además, aplica conocimiento proveniente de campos como lo son administración, ingeniería, comunicaciones, ciencias puras, relaciones humanas, y otras que pueden tener importantes contribuciones para la realización y el estudio de los sistemas de información. Por estas razones se hará brevemente una revisión de los elementos principales de la teoría de sistemas.

Un sistema es un conjunto de variables o componentes, organizados, interactuantes, interdependientes e integrados. Además, un sistema posee objetivos o metas a alcanzar.

## Elementos de un sistema

Los elementos de un sistema son integrados, como se muestra en la figura 1.1

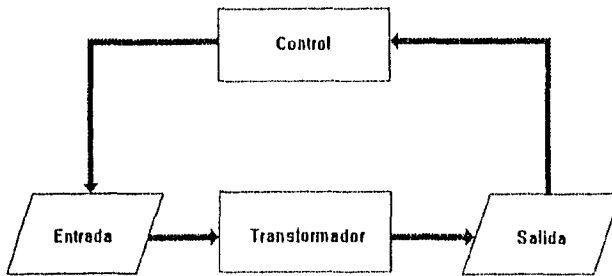


Figura 1.1 Partes componentes de un sistema

El sistema transforma la entrada en salida. Un mecanismo de control monitorea el sistema y regula su operación para que el proceso de transformación sea ejecutado propiamente.

Un sistema debe de tener más de un elemento. Una roca, por ejemplo, no es un sistema. Sin embargo, puede ser parte de uno, como lo puede ser una pared.

¿Qué es un subsistema? Un subsistema es simplemente un sistema que esta contenido dentro de otro. Esto quiere decir que los sistemas existen en diferentes niveles.

¿Qué es un supersistema? Cuando un sistema es parte de un sistema mayor, el sistema mayor es llamado supersistema.

El ambiente es externo al sistema; este posee todo lo que puede estar fuera de control del sistema. Este ambiente externo determina en cierto grado el desempeño del sistema, por lo que el sistema y su ambiente externo están interrelacionados y son interdependientes. Todos los elementos de un sistema deben de tener alguna relación lógica. Los sistemas mecánicos fácilmente satisfacen estos requerimientos. Por ejemplo, un automóvil, ha sido diseñado para realizar una función específica, y todas sus partes funcionan como un sistema sencillo. Mucha veces se asume que los elementos de un sistema deben de funcionar de una manera perfectamente sincronizada. Aunque esto es deseable, no es necesario.

Los recursos son todos los medios disponibles para que el sistema pueda alcanzar los objetivos fijados. En contraste al ambiente externo, los recursos están dentro del sistema y están bajo su control.

Un sistema esta formado por componentes que son tareas, actividades, inisiones, o partes que son ejecutadas con el fin de cumplir un objetivo.

### **Propósito común para alcanzar un objetivo.**

Un sistema es diseñado para alcanzar uno o más objetivos. Todos los elementos trabajan juntos y dirigidos hacia la meta común del sistema en vez de tratar de alcanzar metas por separado. El manejo del sistema consiste en actividades orientadas a la planeación y al control. La planeación debe de abarcar los objetivos fijados, la correcta utilización de los recursos, el desarrollo de planes que abarquen y coordinen las diferentes actividades a realizar, y a definir la estrategia para la convivencia con el ambiente externo. Se debe de llevar un control para la ejecución de los planes. Asociado con este control están el flujo de información y la retroalimentación, con los cuales el sistema podrá evaluar sus planes.

Lo que se muestra a continuación son, lo que algunos teóricos han llamado, las bases de la teoría general de sistemas [LUCAS85]<sup>1</sup>.

1. Los componentes de un sistema están interrelacionados y son interdependientes; los componentes sin relación e independientes no constituyen un sistema. De hecho una de las más importantes tareas del estudio de un sistema es determinar las relaciones entre los componentes.

2. Un sistema es visto como un ente. No es necesario dividirlo en sus partes constitutivas, particularmente si con esto se llegara a perder la perspectiva de este como entidad. En muchos casos la concentración se puede fijar en algún subsistema que constituye un sistema mayor, pero sin ignorar toda la gran estructura que provee el sistema en su totalidad.

3. Los sistemas son buscadores de metas; los componentes interactuantes deben de alcanzar un objetivo o bien un estado final, entonces, una posición de equilibrio y logro debe de obtenerse.

4. Los sistemas tienen entradas y salidas. Los sistemas producen salidas que otros sistemas utilizan como entradas.

5. Los sistemas transforman las entradas en salidas; generalmente la forma que tiene la salida difiere de la forma de la entrada.

6. Los sistemas exhiben entropía. La entropía es un término utilizado en termodinámica, que describe el estado de un sistema cerrado (sin entradas del exterior) en el cual todos los elementos tienden hacia la desorganización y hacia la falta de habilidad para obtener y procesar entradas, por lo que el sistema no puede producir salidas.

7. El sistema debe de tener formas de regular la interacción de sus componentes, de tal forma que los objetivos sean alcanzados. Asociados con esta función reguladora están: la planeación, el control y la retroalimentación.

---

<sup>1</sup> LUCAS, Henry C. The Analysis, Design and Implementation of Information Systems. Mc Graw Hill, 1985

8. Los sistemas generalmente están constituidos por subsistemas más pequeños. Los sistemas de menor tamaño integrados dentro de sistemas mayores crean jerarquía. Esta jerarquía es una característica de la teoría de sistemas.

9. Generalmente existe cierta diferenciación en sistemas complejos; esto es, unidades especializadas desempeñan tareas especializadas.

10. Los sistemas generalmente exhiben un estado final que puede ser alcanzado desde diferentes maneras o desde diferentes puntos de inicio. En otras palabras, existen múltiples formas de alcanzar las metas de un sistema.

### **Análisis y diseño de sistemas**

Las diferentes formas de visualizar el análisis de sistemas se enfocan a un mismo objetivo: entender un sistema complejo y modificarlo de alguna forma. Las modificaciones pueden ser un nuevo subsistema, nuevos componentes, un nuevo conjunto de transformaciones, etc. El objetivo es mejorar el funcionamiento del sistema para hacerlo más eficiente, modificar los objetivos del sistema, cambiar las salidas, alcanzar las misma salida pero con diferente entrada, o realizar alguna función similar.

Normalmente se sigue un cierto procedimiento que incluye lo siguiente :

1. Definir el problema. ¿Por qué nuestro sistema no es lo suficientemente satisfactorio?
2. Entender el sistema y definirlo. Puesto que los sistemas son jerárquicos (subsistemas que existen dentro de otro sistema mayor) y se interrelacionan con sus ambientes externos, puede ser muy difícil definir que es lo que constituye el sistema que se estudia. Esta actividad puede ser subdividida; se harán las siguientes preguntas para entender y desarrollar nuestro sistema:
  - a ¿Cuáles son la variables y componentes, del sistema?
  - b ¿Cómo se interrelacionan estas variables? y ¿Cómo se relacionan con el ambiente externo?
  - c ¿Cuáles son los alcances de el sistema?
3. ¿Qué alternativas existen para alcanzar nuestros objetivos? ¿Qué oportunidad existe de mejorar nuestro sistema? ¿Cuál es el costo? y ¿Puede llegar a ser implementado?
4. Elegir una de las alternativas definidas en el punto anterior.
5. Implantar la alternativa.
6. Si es posible, evaluar el impacto de los cambios que se han hecho al sistema.

De lo anterior expuesto se obtienen los siguientes componentes: El problema, el estudio, la solución y las conclusiones.

## **IMPACTO DE LAS COMPUTADORAS**

La experiencia ha demostrado que la implantación de un sistema puede fallar si el impacto de este no es contemplado. La experiencia, también, revela que el impacto de este puede afectar tanto a los individuos como a las organizaciones.

### **Impacto en la organización**

La empresa, o bien la organización puede volverse a centralizar como resultado de la llegada de la tecnología de las computadoras; la disponibilidad de la información puede impulsar a la gerencia a centralizarse. La tendencia, antes de la llegada de las computadoras era la descentralización, ya que una gerencia centralizada no podía hacer frente a toda la cantidad de información y además, tomar todas las decisiones requeridas dentro de la organización.

Hoy, la computadora brinda la oportunidad de centralizar la toma de decisiones, por lo que la organización puede ser más estrechamente controlada por un pequeño grupo de directivos. En una organización la distribución del poder, es muy importante. El poder es la potencial influencia hacia otros, que nos permite obtener ciertos beneficios. En una organización, diferentes departamentos tienen diferentes niveles de poder. Existen cuatro condiciones que generalmente determinan poder, el departamento de sistemas puede llegar a cumplir estas condiciones y, de esta forma alterar la distribución de poder en la organización.

El primer determinante de poder es la capacidad de disminuir la incertidumbre de otros departamentos. La incertidumbre es la deficiencia de información acerca de eventos futuros; esta deficiencia hace que el éxito de estos eventos sea menos predecible. El departamento de sistemas de cómputo puede llegar a proporcionar información, y uno de los mayores usos de la información puede ser la reducción de la incertidumbre acerca de eventos presentes o futuros.

El segundo determinante es la facilidad con que un departamento puede llegar a ser reemplazado. En realidad no existen muchas alternativas. Una empresa que este a disgusto con su departamento de sistemas lo puede eliminar por completo, pero el caos que generaría la transición a otro grupo podría ser muy costoso. Otra alternativa puede ser buscar un despacho que proporcione el servicio de proceso de datos, pero el reemplazar las aplicaciones puede resultar demasiado costoso y, además ser peligroso. A este servicio de subcontratación de servicios se le llama outsourcing.

De lo anterior se concluye, que no existen muchas posibilidades de reemplazar por completo todo un departamento de sistemas.

Un tercer determinante está definido por la cantidad de relaciones con los otros departamentos. Un número grande de relaciones, hace que su potencial de poder sea grande. El departamento de sistemas puede aceptar el proceso de información proveniente de toda la organización y así, proveer de uno o más servicios a cada uno de los componentes de esta.

El último determinante es el grado de interdependencia con otros departamentos en la organización. Por ejemplo, el departamento de sistemas exhibe una dependencia recíproca con los departamentos usuarios, se dice que son mutuamente dependientes.

De estos cuatro determinantes, el departamento de sistemas posee un alto grado de apoyo en cuanto a la reducción de la incertidumbre, la experiencia revela que este es el determinante más importante hacia la obtención de poder. En el grado en el que el departamento de sistemas desarrolle más y mejores aplicaciones para los diferentes departamentos, crecerá su influencia dentro de la organización.

### **Impacto en los individuos**

Muchos de los primeros estudios acerca del impacto de las computadoras se enfocan a la reacción de usuarios individuales. Algunos de estos se refieren a aspectos psicológicos, otros se enfocaban a los niveles de empleo. Aunque no es fácil hallar una tendencia, se puede decir que el impacto de las computadoras en los niveles de empleo no ha sido más grande que el que pudo haber causado otro cambio tecnológico.

En adición a los niveles de empleo, se puede hablar de impacto en el desarrollo de trabajos y a la estructura interna de estos. Se sugiere que las computadoras realicen más actividades innovadoras dirigidas a puestos directivos y encaminen a la organización a poseer una élite directiva, por ejemplo. Esto puede acentuar los espacios entre niveles de toma de decisiones dentro de la organización, y en el peor de los casos a la desaparición de estos.

Con los individuos el problema real puede llegar cuando entran en conflicto el departamento de sistemas y el departamento de usuarios. El conflicto puede ser causado por varias condiciones, una de ellas puede ser la transferencia de poder de los usuarios a sistemas. Si el sistema no cumple con las expectativas del usuario, ya sea en cuanto a implantación o a compromisos programados, el conflicto se hará presente.

Existen algunas condiciones que poseen cierto potencial para crear conflicto. La primera de estas condiciones, puede ser la gran dependencia entre los involucrados, ya que la falla de uno de ellos puede causar serios problemas al otro. La diferencia de tareas puede también crear conflicto. El trabajo relacionado con sistemas es altamente especializado y existen muchas diferencias entre las tareas de un programador, un analista, un gerente y el usuario final. Por lo que la comunicación entre ellos podría encontrar diversos obstáculos y, finalmente fallar. Ocasionalmente, cuando la gente depende de recursos comunes, los conflictos pueden surgir. Por ejemplo, varios departamentos pueden estar compitiendo por recursos de cómputo.

No todas las condiciones anteriores forzosamente existen en una organización. Sin embargo, la relación entre el departamento de sistemas y los usuarios tienen el potencial de crear conflictos.

Todo lo anterior no excluye a ningún nivel jerárquico dentro de una organización.

## EL SOFTWARE

Hace veinte años, el término software era prácticamente desconocido por la gran mayoría de los integrantes de la comunidad profesional. Hoy en día, el término está mucho más difundido.

Pero, ¿Qué se entiende por software? Software: (1) Instrucciones (programas de computadora), que cuando son ejecutados proveen una función deseada, (2) Estructuras de datos que permiten a los programas manipular adecuadamente la información.

### Características del Software

Para poder entender un mejor que es lo que significa el software, es importante examinar las características de este, que lo hacen diferente de otros componentes del sistema.

Cuando es construido el hardware, el proceso creativo es traducido a su forma física.

El software es lógico, mas no físico. Entonces, el software posee características que son considerablemente diferentes que las del hardware. El software es desarrollado, el hardware es manufacturado. Aunque existen algunas semejanzas entre el desarrollo y la manufactura, estas dos actividades son fundamentalmente diferentes. En ambas actividades la alta calidad se logra mediante el buen diseño. Los costos del software se concentran en el diseño. Esto significa que los proyectos no pueden ser manejados como si fueran de manufactura. El software no se gasta. En el hardware, la curva de fallas de este muestra que la alta incidencia de estas es a temprana edad. Estas fallas son generalmente atribuidas a defectos de diseño o de manufactura. Cuando los defectos son corregidos la tasa de errores disminuye, y se mantiene en un nivel bajo por un determinado periodo de tiempo. A medida que el tiempo pasa, esta curva vuelve a ascender a causa del uso de los componentes, del polvo, de la vibración, las variaciones de temperatura, y de otros factores. Simplemente se gasta.

El software no es susceptible de ser afectado por los factores ambientales. En teoría la curva de fallas para este debería ser mucho más estable. Esto es, los errores se presentarán en las etapas tempranas del programa y causarán alta incidencia de fallas. Sin embargo, una vez que estas hallan sido corregidas, asumiendo que no se introducirán otros errores, la curva se aplanará. Sin embargo, la realidad es otra: El software no se gasta, se deteriora.

Esto al parecer es una contradicción. Si se considera que durante la vida del software, este ha sufrido cambios, mantenimiento. Así como los cambios van siendo realizados, errores potenciales podrían ser introducidos al software, causando que la curva de errores presente picos de considerable impacto. Antes de que la curva pueda regresar al estado en que ocurrió la anterior falla, otro cambio fue necesario, causando que la curva vuelva a presentar un pico. Lentamente el nivel de fallas empieza a elevarse; el software, debido a los cambios, empieza a deteriorarse.

Un aspecto que ilustra bien la diferencia entre el hardware y el software es que cuando un componente de hardware se gasta, este es remplazado por una parte de repuesto. Para el software



no existen estas partes de repuesto. Cada falla del software indica un error en el diseño o en el proceso mediante el cual el diseño fue traducido a código de máquina ejecutable. Por esto, el mantenimiento de software implica mas complejidad que la del mantenimiento del hardware

### Proceso del software o Componentes del software

El software para computadora es información que puede existir en alguna de dos formas básicas: Componentes no ejecutables por máquina y componentes ejecutables por máquina

La manera en que el software es traducido de forma no ejecutable a ejecutable, es expresada por la figura 1.2.

El diseño del software es traducido a lenguaje de programación. Este código es procesado por un traductor que lo convierte en sentencias ejecutables por la máquina

Idealmente la gente se debería de comunicar con computadoras por medio de su lenguaje natural. Desafortunadamente, el tamaño de los vocabularios, la dificultad de las gramáticas y el uso de contextos han impedido la comunicación de hombre máquina, por medio del lenguaje natural. Todos los lenguajes de programación son lenguajes artificiales. Cada uno tiene vocabulario limitado, una gramática explícitamente definida y ciertas reglas de sintaxis, así como de semántica. Es posible diferenciar a los lenguajes en lenguaje de máquina, lenguaje ensamblador, lenguaje de alto nivel y lenguajes no procedurales. Los tres primeros son conocidos como las tres primeras generaciones de lenguajes para computadora. En estas, el programador debe de ocuparse de tanto de la estructura de la información como del control del programa por si mismo. De aquí que se les llame lenguajes procedurales.

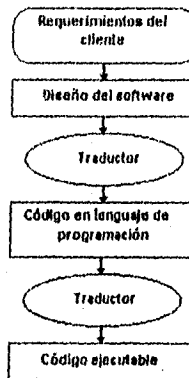


Figura 1.2 Pasos en la traducción del software

Durante la década de los ochentas, una cuarta generación de lenguajes comenzó a surgir. Estos lenguajes fueron llamados no procedurales. En vez de requerir que el desarrollador especificara el procedimiento en detalle, en el lenguaje no procedural especifica el resultado deseado, mas no especifica la acción requerida para lograr el resultado. Hasta la fecha los lenguajes de cuarta generación (Fourth Generation Languages, 4GLs) han venido siendo utilizados principalmente en aplicaciones orientadas a base de datos.

### **Ingeniería de sistemas de computadora**

Durante las dos últimas décadas, los sistemas basados en computadora han introducido un nuevo orden.

La ingeniería de software y la ingeniería de hardware caen dentro del vasto campo de la ingeniería de sistemas de computadora. Cada una de estas dos disciplinas representan el intento por tener un orden en el desarrollo de sistemas basados en computadora. Las técnicas ingenieriles para el hardware ha crecido con el diseño electrónico y ha alcanzado una relativa madurez. Las técnicas de diseño están bien establecidos, los métodos de manufactura son continuamente mejorados y son de alta confiabilidad.

Por el lado del software, este ha reemplazado a el hardware en ser el componente más difícil de planear, y implantar exitosamente, y más peligroso de manipular. La demanda por software continua creciendo en número, complejidad y en nuevas aplicaciones.

La utilización de las técnicas para producir software basados en conceptos de ingeniería, apenas ha empezado a ganar una aceptación considerable. Por fortuna los tiempos están cambiando, de ser considerado una especie de arte, el desarrollo de software empieza a verse como toda una disciplina basada y apoyada en la ingeniería.

### **Sistemas basados en computadora**

La palabra sistema es posiblemente el término mas utilizado en el léxico técnico. Se define un sistema basado en computadora como el conjunto de elementos organizados con el fin de lograr un objetivo mediante el proceso de información

Los elementos que constituyen un sistema de información son los siguientes:

- **Software.** Programas de hechos en algún lenguaje de computadora, así como las tecnologías utilizadas en esta tarea, por ejemplo estructuras de datos, programación orientada a objeto, técnicas de desarrollo e implantación, etc.
- **Hardware.** Dispositivos electrónicos que proveen la capacidad de cómputo, y dispositivos electromecánicos que proveen servicios hacia y del exterior.
- **Personal.** Individuos que realizan alguna o algunas de las siguientes funciones: Operar, diseñar, desarrollar, implantar o utilizar del hardware y/o el software.

- Bases de datos. Son grandes colecciones de información, organizadas y que son accedidas vía software.
- Documentación. Manuales e información descriptiva que es útil e indispensable para la utilización y operación del sistema.
- Procedimientos. Son los pasos que definen el uso específico de cada elemento del sistema.

Una característica de los sistemas basados en computadora es que los elementos que componen un sistema puede también representar un macro elemento que a su vez represente por sí otro sistema.

La ingeniería sistemas de computadora es una disciplina orientada al desarrollo de software de alta calidad. Es una actividad solucionadora de problemas. Comienza con la identificación de los problemas o bien de las actividades susceptibles de ser implantadas en sistemas de computadora. Y termina con la presentación de un sistema que resuelve o simplifica la actividad analizada.

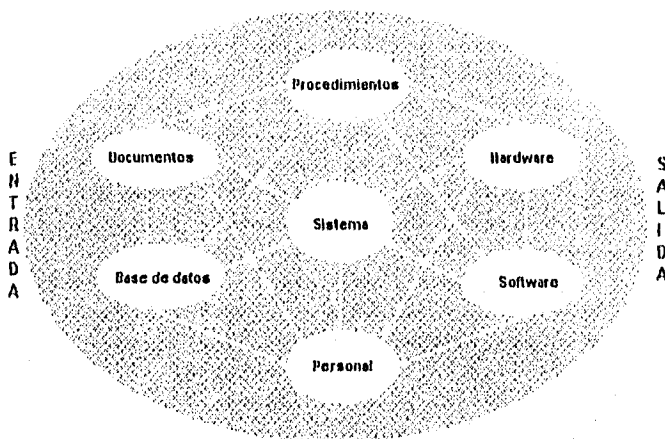


Figura 1.3 Elementos de un sistema basado en computadora

### El elemento software

El llamado elemento software de un sistema basado en computadora esta compuesto por software de aplicación y por software del sistema. El software de aplicación implanta el procedimiento requerido para adecuar las funciones del proceso de información. El software del sistema implanta las funciones de control que permiten a el software de aplicación interactuar con otros elementos del sistema.

## Aplicaciones del Software

Sin importar el área de aplicación de un sistema basado en computadora, este puede ser representado utilizando un modelo entrada-proceso-salida.

El software suele ser utilizado para recibir información de entrada, la cual es generalmente provista ya sea por alguna por alguna fuente externa o, por otro elemento del sistema.

Cuando el sistema basado en computadora requiere una interfaz hombre-máquina interactiva, el software implanta una conversación de entrada-salida. Los procedimientos de entrada de datos son implantados por medio de software, desplegando la información en forma gráfica. Cuando la información de entrada debe ser adquirida desde un dispositivo, el software maneja el hardware para que esto sea posible. Finalmente el software es el encargado de establecer la comunicación con los sistemas de base de datos. El software implantado procesa los algoritmos necesarios para alcanzar los objetivos fijados. En general el proceso de los algoritmos transforma los datos de entrada en información o bien datos que pudieran servir a otro elemento del sistema. Hoy en día los mas comunes tipos de proceso son numéricos y no numéricos; en los cuales todos los pasos y condiciones están predefinidas. Sin embargo, surge una nueva categoría de proceso de algoritmos llamado sistema experto. A diferencia de los algoritmos convencionales el sistema experto hace uso de hechos especificados y reglas de inferencia, lo cual permite al software exhibir en un número limitado de problemas habilidades parecidas a las humanas.

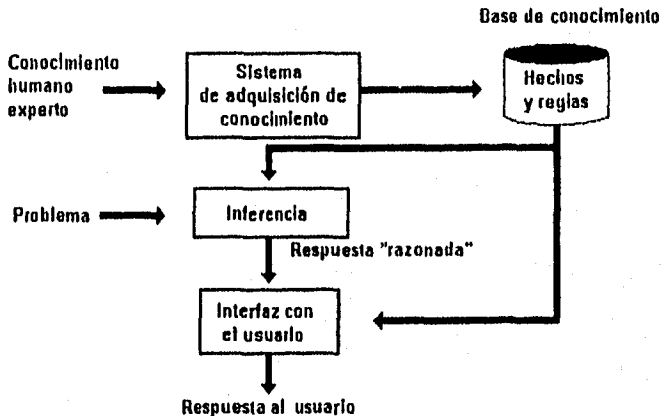


Figura 1.4 Estructura clásica de un sistema experto

Siendo práctico en el uso del software, este debe de producir como salida información útil ya sea para otro elemento del sistema o bien para alguna parte de la organización.

## **Importancia del software**

En la actualidad, el software ha sobrepasado al hardware como elemento clave para el éxito de los negocios, productos y sistemas. En esencia el software es, a menudo el factor que inclina la balanza.

Durante las primeras tres décadas de la era de la computación, el reto era desarrollar hardware que redujera el costo de procesar y almacenar datos. Durante las décadas de los 80's y los 90's, los avances en la microelectrónica han redundado en la disminución de los costos e incremento del poder de computo

Hoy en día el problema es diferente. El reto primario es reducir los costos y mejorar la calidad de las soluciones basadas en computadora; soluciones que son implantadas con software. El contexto en el cual el software ha sido desarrollado esta íntimamente ligado a la evolución de las computadoras. El mejor desempeño de el hardware, la reducción de las dimensiones físicas y la reducción de costos ha ocasionado que los sistemas se hayan vuelto mas sofisticados.

La evolución del software dentro del contexto de las áreas de aplicación para los sistemas basados en computadora ha sido como sigue: Durante los primeros años de los sistemas desarrollados para computadora, el hardware sufrió cambios continuamente mientras el software fue algo secundario y con menos importancia. El software era algo poco controlado, a menos que los costos se elevaran o bien los compromisos contraídos no se fueran a cumplir. Durante este periodo la orientación batch fue usada para la mayoría de los sistemas. Notables excepciones fueron los sistemas interactivos de las líneas aéreas, así como los de tiempo real de las secretarías de defensa. Pero en la mayoría, sin embargo, el hardware estaba dedicado a la ejecución de un solo programa, el cual a su vez estaba dedicado a una aplicación en específico. Durante esos años el hardware que predominaba era el orientado a propósito general. El software, por otro lado, era particularmente diseñado y tenía una distribución muy limitada.

Gracias a todos esos años se ha aprendido mucho acerca de la implantación de sistemas basados en computadora, pero poco relativamente acerca de ingeniería de software.

La segunda era de la evolución de los sistemas basados en computadora. Esta era comienza en los principios de los años sesentas y llega hasta fines de los setentas. La multiprogramación y los sistemas multiusuarios introdujeron nuevos conceptos de interacción hombre-máquina. Las técnicas interactivas abrieron un nuevo mundo de aplicaciones y nuevos niveles de sofisticación tanto de software como de hardware. Los sistemas de tiempo real son útiles para coleccionar, analizar y transformar datos provenientes de múltiples fuentes, con lo cual se pueden controlar procesos y producir una salida útil en milisegundos, en vez de minutos. Los avances en los dispositivos de almacenamiento llevan a el surgimiento de la primera generación de los sistemas manejadores de bases de datos.

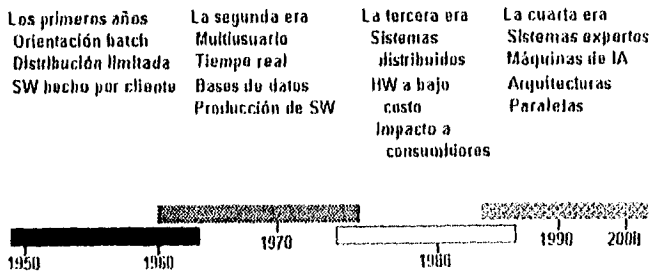


Figura 1.5 Evolución del software

La segunda era también se caracteriza por el uso de software ya como un producto, por lo que aparecen las casas de software. El software es desarrollado para muchos y diferentes mercados. Se crean programas para minis y para mainframes. La industria, las instituciones gubernamentales y los académicos se empiezan a enfrentar por la supremacía y la vanguardia, esto obviamente para atraer dinero. El número de sistemas basados en computadora comenzó a crecer, las publicaciones acerca de software comenzaron a surgir. El desarrollo casero produjo cantidades considerables de programas. Pero, un gran problema se veía venir: La gran mayoría de estos programas tuvieron que corregir aquellas fallas y errores que iban siendo detectadas, también, estos programas eran modificados a requerimiento del usuario o bien adecuado por la necesidad de algún nuevo hardware adquirido. Estas actividades fueron llamadas, mantenimiento de software. El esfuerzo y los recursos dedicados a el mantenimiento del software crecían en forma alarmante. Pero resultaba que la propia naturaleza de los programas los hacían imposibles de mantener. Una crisis del software había llegado.

La tercera era de la evolución de las computadoras comenzó a mitad de la década de los setentas y terminó a finales de los ochentas. Caracterizaron a esta época los sistemas distribuidos, las computadoras con varios procesadores, las redes, la comunicación digital y el proceso en línea.

La tercera era es también caracterizada por el advenimiento y amplio uso de las computadoras personales. Los microprocesadores también son utilizados en una gran variedad de aparatos, enseres, máquinas y equipos. La computadora personal ha sido el catalizador para el crecimiento de muchas compañías de software, así como para la proliferación y uso común de la computadora. Mucha gente gastó más en adquirir software para su computadora personal que en la computadora misma.

La cuarta era de las computadoras comenzó en la parte final de los ochentas y todavía no termina. Se avcena una quinta generación de computadoras y el software relacionado tendrá un

alto impacto en la industria en general en todo el mundo. Las técnicas de la cuarta generación (Fourth Generation Techniques, 4GT) han cambiado la manera en que trabaja la comunidad que desarrolla software. La inteligencia artificial y los sistemas expertos finalmente se han movido de los laboratorios a las aplicaciones prácticas.

A pesar de estar ya en una cuarta era de la evolución del software, la llamada crisis del software continúa intensificándose. Es posible describir la crisis del software en los siguientes tres puntos:

1. La sofisticación del hardware ha sobrepasado nuestra habilidad de crear software que pueda aprovechar el potencial de este.
2. La habilidad para crear nuevos programas no ha podido con la demanda.
3. La habilidad para mantener los programas existentes es amenazada por la falta de recursos y la pobreza del diseño.

Como respuesta a la crisis del software, la práctica de la ingeniería de software ha tenido que ser adoptada por la industria. La exigencia de los usuarios y la profesionalización de los desarrolladores ha redundado en el establecimiento de una comunidad informática cada vez más orientada hacia la calidad. □

# 2

---

## *Sistemas de Información*

### INTRODUCCIÓN

□ La información es un recurso crítico y fundamental de las organizaciones. Es la liga indispensable que mantiene a los componentes de una organización unidos para una mejor operación y coordinación, además de apoyar la supervivencia de la misma en un ambiente competitivo.

La información son aquellos datos que se han puesto en un medio entendible y que se convierten en un elemento útil para la toma de decisiones. La información incluye la comunicación y recepción de inteligencia o conocimiento. Esta, es utilizada para notificar, estimular, reducir la incertidumbre, revelar alternativas adicionales, o crear influencia. Y, especialmente en los negocios, la información puede llegar a pronosticar.

La información consiste de datos, imágenes, texto, documentos, y/o voz. A menudo intrínsecamente relacionada, pero siempre organizada de una manera entendible.

En muchas ocasiones el término datos es utilizado para generalizar el de información, pero es importante recordar que la información es mucho más que solo números.

Un aspecto que se debe de recordar es el ciclo de la información. Los datos son procesados a través de modelos para crear información. Cuando la información llega a el destinatario este podrá apoyarse en esta para tomar alguna decisión y realizar una acción; esto a su vez crea otras acciones o eventos, los cuales generarán un número indeterminado de datos dispersos los cuales serán capturados y servirán como entrada; y el ciclo comienza de nuevo.

La calidad de la información se sustenta en tres aspectos: precisión, oportunidad y relevancia.



La precisión significa que la información debe de estar libre de influencias y distorsiones. Además de estar libre de errores, esta debe de ser clara y debe de reflejar el significado de los datos en que esta basada.

La oportunidad significa, estar en el lugar y en el momento adecuado. Y, en muchos casos, inclusive, con anticipación.

La relevancia es el grado de utilidad en una situación determinada. Además, se debe de tener en cuenta que cierta información puede ser relevante para algunos y para otros no.

### **Sistemas de información**

¿Qué es un sistema de información? Como ya se revisó en la Introducción, se define un sistema de información como un conjunto de procedimientos organizados que, cuando son ejecutados, proveen de información útil para la toma de decisiones y/o para el correcto control de una organización. La información es una entidad tangible o intangible que reduce la incertidumbre acerca de un estado o evento. Pero, el nombre de sistema de información ha sufrido una serie de cambios que refleja la evolución de la disciplina. Originalmente el término fue utilizado como indicador de sistematización electrónica en una organización. Se le llamó proceso de datos automático (Automatic Data Processing, ADP), proceso de datos electrónico (Electronic Data Processing, EDP), o solo proceso de datos (Data Processing, DP). Sin embargo, cada una de los anteriores sugería un muy limitado rol dentro de la organización, confiriéndole importancia puramente técnica y con un cierto reconocimiento por incursionar en áreas operativas y, últimamente, de organización y estrategia.

El término información ha comenzado a ser utilizado como un preciso descriptor de lo que esta sucediendo en la práctica. De manera similar, el término sistemas empezó a asociarse a el software, la práctica, la gente y los procesos relacionados. De aquí, la frase sistemas de información fue acuñada para describir aquellos recursos que forman parte de la organización y que son responsables del manejo de la información dentro y fuera esta, así como de proporcionar a los otros sistemas de la organización la información que les sea útil para lograr los objetivos previstos.

Como se ha visto un sistema de información es una función de servicio interno. Su función primaria es asistir a otras unidades dentro de la organización para que puedan realizar sus funciones de una manera mas efectiva y eficiente. A menos que los sistemas de información sean vendidos como servicio a los usuarios finales, estos no producirán un servicio externo o bien un ingreso.

Es de máxima importancia que la persona al frente de el sistema de información tenga un excelente conocimiento de los requerimientos de su cliente. Sin un conocimiento profundo de la

operación del cliente, es difícil que el sistema de información entregue servicios que sean consistentes con los objetivos planteados.

Lo anteriormente expuesto parece se obvio, pero una de las tareas más difíciles para crear un sistema de información es la de llegar a un entendimiento de acerca de las necesidades del cliente. Muy a menudo, el encargado del sistema de información está ansioso por implantar una solución de alta calidad técnica sin haber entendido por completo el problema del cliente; esto resulta en un cliente insatisfecho. El responsable de los sistemas de información tiene generalmente un amplio conocimiento de los aspectos técnicos y el cliente tiene un profundo conocimiento de la operación y funcionamiento de su especialidad. Solo con la cooperación de estas dos entidades será posible el diseño de un sistema de información que en realidad cumpla con las necesidades del cliente.

Como se ha visto las decisiones que involucran a un sistema de información son dependientes de las metas y objetivos fijados ya sea por un individuo, para varios departamentos, o bien para toda una organización. De hecho, los sistemas de información deben actuar como un consultor técnico para la toma de decisiones.

## **EVOLUCIÓN DE LOS SISTEMAS DE INFORMACIÓN**

Los esfuerzos iniciales para aplicar la computadora en el área de los negocios se enfocaron a los datos. Hoy, la comunicación y la consulta reciben la mayor atención.

### **Los datos, foco inicial de atención**

Durante la era previa a las computadoras, las organizaciones generalmente ignoraban que la información era necesitada por los niveles de mando. Esta práctica continuó con las primeras computadoras, estas fueron restringidas al uso de propósitos únicamente contables. Durante este periodo, el cual llegó hasta mediados de los sesentas, las aplicaciones fueron llamadas proceso electrónico de datos, EDP. Este término es ahora difícilmente utilizado. EDP implicaba limitar el uso de las computadoras para proceso de contabilidad de datos en vez de producir información útil para el manejo de los sistemas o de la organización.

El término que se utiliza para describir esta aplicación primaria de la computadora, la cual continua, es el proceso de datos, DP. El DP produce cierta información, aunque esta no sea su misión principal.

### **El nuevo foco de atención de la información**

Fueron algunos perspicaces individuos quienes se dieron cuenta que la computadora podía hacer más que solo procesar datos. Fue en IBM donde se desarrolló un uso para las computadoras conocido como recuperación de información (Information Retrieve, IR). Este desarrollo ocurrió a

finales de los cincuentas y principio de los sesentas. La recuperación de información no calculaba ni acumulaba datos. Esta estaba dedicada a el almacenamiento de datos especializados, como lo son títulos y sumarios de publicaciones impresas o bien ciertos expedientes.

El IR fue el primer paso hacia la utilización de la computadora como un sistema de información.

En 1964 una nueva generación de equipo de cómputo fue introducido. Este extendió una fuerte influencia en la forma en que este era empleado. Las nuevas computadoras fueron las primeras en utilizar microcircuitaria de silicio, y ellas ofrecían la oportunidad de mayor poder de proceso por dólar. Una compañía podía adquirir una computadora rápida con unidades de gran capacidad de almacenamiento y equipo de comunicación de datos por un, relativamente, bajo costo. El concepto de manejo de sistemas de información, MIS, fue promovido por fabricantes de computadoras para justificar equipo adicional. El concepto MIS reconoció que las aplicaciones debían ser implantadas bajo el propósito primario de producir información útil a los que se encargaban de la toma de decisiones. El concepto fue rápidamente adoptado por muchas de las grandes organizaciones puesto que en realidad si existía la necesidad de un mejor manejo de la información.

El trecho caminado por estos pioneros no fue fácil. Existen varias razones por las cuales muchos de los primeros esfuerzos fallaron. Una generalizada falta de literatura acerca de computadoras entre los usuarios, una ignorancia acerca de el manejo de los roles entre los especialistas, el equipo de cómputo era caro y limitado, etc. Pero un error en particular caracterizó a los primeros sistemas, eran demasiado ambiguos. Las grandes organizaciones creían que podían crear sistemas de información gigantes para poder dar soporte a todos los gerentes. Los diseños de los sistemas crecieron y crecieron, hasta que se volvieron imposibles de manejar. Algunas no desistieron, siguieron invirtiendo recursos, y eventualmente desarrollaron sistemas funcionales, aunque de tamaño más modesto del originalmente proyectado. Otras organizaciones decidieron terminar por completo con la idea de MIS y regresaron al concepto de EDP.

### **El interés por el soporte a decisiones**

Mientras algunos veían como las organizaciones lidiaban con sus gigantescos MIS, algunos científicos de el Massachusetts Institute of Technology (MIT) formularon una idea diferente. Estos científicos llamaron a su concepto sistemas de soporte a decisiones (Decision Support Systems, DSS). Un DSS es un sistema productor de información enfocado a un problema en particular que algún gerente debe de resolver, para que después cierta(s) decisión(es) sean tomadas.

En vez de tratar de implantar un MIS gigante, los que impulsaban los DSS recomendaban enfocarse en problemas separados y diseñar mas de un DSS para cada problema.

Un DSS puede ser aplicado con mayor éxito a los problemas semiestructurados. Un problema estructurado es aquel que posee elementos conocidos y cuyas relaciones son entendidas. Un problema no estructurado es justamente lo opuesto, ni los elementos ni las relaciones son conocidas. Entonces, un problema semiestructurado consiste en que algunos elementos y algunas relaciones son conocidas y entendidas, y otras no.

En la actualidad existe una corriente bajo la cual se busca aplicar la inteligencia artificial (Artificial Intelligence, AI), a problemas de negocios. La idea básica de la AI es que la computadora pueda ser programada para ejecutar algo de la lógica del razonamiento humano. Una subclase de la AI, son los sistemas expertos (Expert Systems, ES), los cuales están recibiendo la mayor parte de la atención. Un Expert System es aquel que desempeña las funciones de un especialista de determinada área. En los próximos años se puede esperar que los sistemas expertos jueguen un rol de creciente importancia.

## **SISTEMAS DE APOYO A DECISIONES**

En los primeros años de los setentas, ciertas compañías comenzaron a producir sistemas de información que se diferenciaban un poco de los sistemas MIS. Estos nuevos sistemas eran menores en términos de tamaño y costo. Estos también eran interactivos, diseñados para ayudar a los usuarios finales a utilizar datos y modelos para así discutir y decidir, más no para resolver, problemas no estructurados y semiestructurados. Cerca de finales de los 80's, estos esfuerzos para asistir decisiones individuales fueron extendidos a grupos y organizaciones enteras.

### **Definición de un DSS**

Las características antes mencionadas pueden servir como una definición de sistema de apoyo a decisiones (DSS): sistemas interactivos bajo el control del usuario, provistos de datos y modelos que servirán como base para discutir y tomar decisiones de acerca de problemas no estructurados o semiestructurados.

La Figura 2.1 muestra un diagrama de un DSS. Un DSS es un sistema integrado, enfocado en decisiones específicas que combinan datos, modelos y software amigable e interactivo, que lo convierten en un sencillo y poderoso sistema bajo el control del usuario.

### **¿Qué es un DSS?**

El concepto de DSS tiene sus comienzos con las opciones de tiempo compartido, a finales de los sesentas. Por primera vez, una persona podía interactuar directamente con una computadora sin tener que pedir ayuda a algún especialista.

La opción de tiempo compartido abrió nuevas oportunidades para el uso de las computadoras.

Con estos antecedentes del entorno de los DSS, se puede definir un DSS, como un sistema que apoya en la solución de problemas semiestructurados proveyendo información o sugerencias concernientes a decisiones específicas. La información que se provee como resultado puede ser reportes, resultados provenientes de modelos matemáticos, y alternativas a analizar para después emitir un juicio o decisión definitiva.

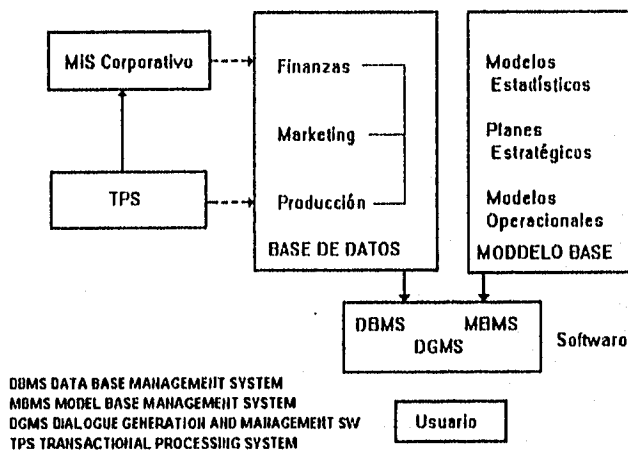


Figura 2.1 Componentes de un Decision Support System

### Objetivo de un DSS

Los objetivos que un DSS debe de alcanzar son:

1. Asistir en la toma de decisiones para solucionar problemas semiestructurados.
2. Apoyar, mas no reemplazar, a quienes toman las decisiones a emitir un juicio.
3. Mejorar la efectividad, mas no la eficiencia, de quienes toman las decisiones.

Estos objetivos relacionados son los tres principios fundamentales del concepto de DSS, estructura del problema, apoyo a decisiones, y efectividad de la decisión.

### **Características de un DSS**

Puesto que un DSS promete apoyar la toma de decisiones, es importante el entender las características únicas de un DSS.

Los DSS están diseñados para soportar el análisis problemas semiestructurados y no estructurados.

Los problemas estructurados son repetitivos y rutinarios, para los cuales algunos algoritmos ya conocidos proveen soluciones. Los problemas no estructurados son novedosos y no rutinarios, para los cuales no existe un algoritmo para solucionarlos. Se puede discutir y decidir acerca de problemas no estructurados, pero no pueden ser resueltos en el sentido de encontrar una respuesta a una ecuación. Los problemas semiestructurados caen en un término medio entre los problemas estructurados y no estructurados.

### **Origen de los DSS**

Es difícil de entender la significancia de un DSS sin considerar donde encajan en la evolución de la tecnología de información y en la arquitectura de las organizaciones. Un DSS es la evolución natural que comenzó con el EDP y el primer sistema de proceso de transacciones y culminó a finales de los sesentas con el MIS y el primer sistema manejador de reportes.

Ya por los setentas, la insatisfacción por el MIS fue emergiendo en la comunidad de usuarios finales. La habilidad de MIS para apoyar la toma de decisiones directamente estaba limitada. Los sistemas MIS no eran interactivos y requerían de la intervención directa de los analistas y los diseñadores para obtener cualquier tipo de información.

Los sistemas eran demasiado lentos como para permitir que el uso de información relevante estuviera disponible día a día. En vez de apoyar la toma de decisiones, los MIS no eran mas que solo generadores de reportes. Además, los avances en la tecnología de mediados de los setentas había hecho posible que los diversos tipos de sistemas pudieran ser desarrollados. Con las minicomputadoras el poder de proceso pudo ser distribuido a los usuarios. Los sistemas operativos permitieron sesiones interactivas y de tiempo compartido. El software más amigable estaba siendo desarrollado, especialmente generadores de reportes y paquetes estadísticos. Las primeras bases de datos y los lenguajes de cuarta generación comenzaban a surgir. De similar importancia fue el desarrollo de el software gráfico, así como el hardware capaz de brindar oportunidades y alternativas para el creciente número de usuarios.

Con la llegada de las microcomputadoras en los ochentas, y de las estaciones de trabajo durante estos años. Las ideas originales de los DSS pudieron finalmente realizarse. Con la nueva arquitectura de información, al menos es técnicamente posible proveer a los usuarios finales con el acceso a grandes bases de datos, modelos analíticos e interfaces amigables.

En suma, la posibilidad técnica existe para proveer el apoyo a toma de decisiones a grupos o a enteras organizaciones. Estos avances técnicos revolucionaron las expectativas de los usuarios finales. Los reportes batch generados por los departamentos MIS simplemente no pudieron con las imágenes presentadas en terminales interactivas bajo el control de usuarios que analizaban datos corporativos con la ayuda de poderosos modelos estadísticos.

Así, fue durante este periodo que se llevó a cabo una activa reconsideración de las organizaciones dedicadas a la toma de decisiones. Existió una creciente falta de conciencia de que muchas decisiones a las que se enfrentaban los gerentes comunes nunca se apoyaban en a las técnicas de la investigación de operaciones.

Los DSS a menudo tienen diferente posición organizacional que los MIS. Los Grupos de DSS a menudo surgen como entidades separadas en divisiones de usuarios finales o bien como un grupo bajo el control de la alta dirección.

Hoy algunas de las ideas de los DSS son comunes; están integrados dentro de MIS existentes y forma parte básica de desarrollos de software de toma de decisiones. De hecho, uno de los resultados del desarrollo de los DSS es una mejora de los MIS. Si se toman en cuenta los factores del ambiente externo y la mayor capacidad para aprovechar los modelos matemáticos fueran agregados a un MIS, este comenzaría a tomar forma de DSS. Puesto que muchos conceptos de DSS han venido siendo agregados a MIS, algunas veces la distinción entre las dos clases de sistemas se convierte en una tarea difícil de realizar.

### **Tipos de decisiones según Simon**

En su libro de 1977 Simon [SIMON77]<sup>2</sup> elaboró en términos de decisiones programadas y no programadas. Aunque él no se dio el crédito de acuñar estos términos, él los describe de manera que han sido ampliamente aceptados para quienes estudian la toma de decisiones.

Las decisiones programadas son repetitivas y con una rutina definida, al grado de que existe un procedimiento definido el cual se ha creado para manejarlo de tal manera que no es necesario tratarlo como un problema nuevo cada vez que este ocurre. Las decisiones no programadas son nuevas, sin estructura y poco consecuentes. No existe un método estrictamente definido para manejar el problema puesto que ninguno exactamente igual a este se ha presentado antes, o porque su naturaleza y estructura son difíciles y complejas, o bien por que su importancia merece un tratamiento especial.

Simon explica que los dos tipos de decisiones son solo los extremos blanco y el negro, y que el mundo era mas bien gris. Sin embargo el concepto de decisiones programadas y no

---

<sup>2</sup> Simon, H. Administrative Behavior. McMillan, 1977

programadas (estructuradas y no estructuradas) es importante puesto que cada uno de ellos hace referencia a técnicas diferentes.

### Fases de la toma de decisiones de Simon

Otra de las contribuciones de Simon es su descripción de las cuatro fases de se siguen para resolver un problema.

Las fases de Simon son:

1. Actividad de inteligencia. Búsqueda de el ambiente propicio para la solución.
2. Actividad de diseño. Invención, desarrollo, y análisis de los posibles cursos de acción.
3. Actividad de elección-selección de un particular curso de acción de aquellos que están disponibles.
4. Actividad de revisión. Evaluación posterior de opciones y seguimiento de resultados.

La Figura 2.2 muestra como las cuatro fases anteriores se relacionan a los pasos de análisis de sistemas. Su actividad de inteligencia se relaciona con los pasos de analizar las partes del sistema en secuencia y procediendo de un sistema a un nivel de subsistema en búsqueda de información del problema. Su actividad de diseño corresponde a nuestro paso de identificar y evaluar alternativas. Y su actividad de elección se relaciona con la de seleccionar la mejor alternativa. Finalmente, su actividad de revisión se relaciona con nuestro paso de implantación de solución y seguimiento.

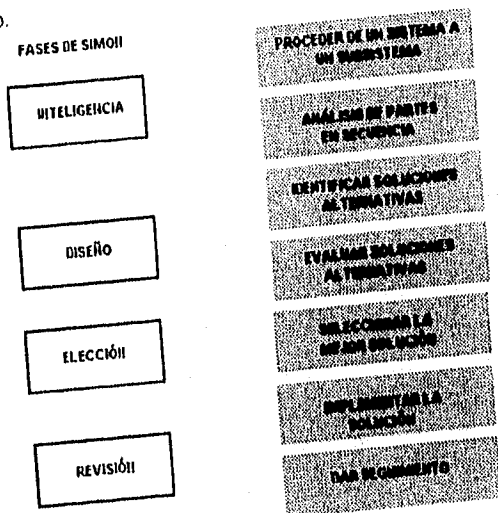


Figura 2.2 Fases para la solución de problemas según Simon



Las fases de Simon, entonces, son otra interpretación del enfoque de sistemas

Un problema completamente estructurado es aquel en el cual las tres fases de Simon - inteligencia, diseño y decisión- son estructuradas. Es decir, es posible especificar algoritmos, o reglas de decisión que permiten que el problema sea identificado y entendido, alternar soluciones para que sean identificadas y evaluadas, y así sea seleccionada una de las soluciones. Un problema no-estructurado, por otro lado, es aquel en el que ninguna de las tres fases esta estructurada. Un problema semiestructurado es aquel en el que una o dos de las fases no esta estructurada.

### Niveles de toma de decisiones

Los niveles son de toma de decisiones son:

La toma de decisiones estratégica es referente a la decisión sobre los objetivos, recursos, y políticas de la organización. Un problema mayor de la toma de decisiones es poder predecir el comportamiento de políticas o eventos de la organización y el medio ambiente, así como hacer coincidir las características de la empresa con el medio ambiente.

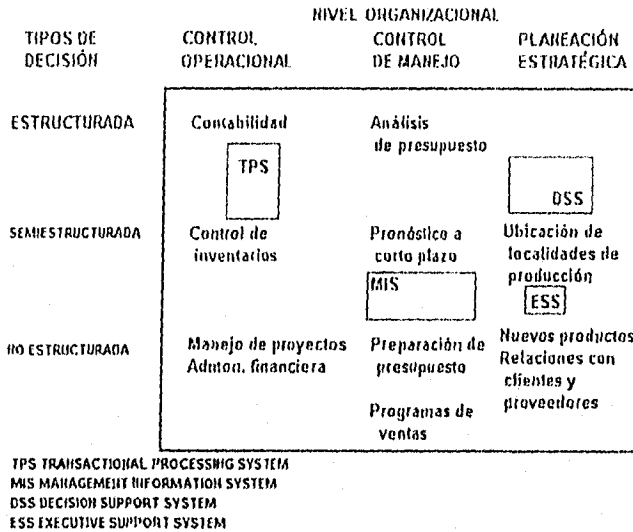


Figura 2.3 Algunos tipos de sistemas basados en computadora

La toma de decisiones para el control gerencial la cual concierne a que tan eficiente y efectivamente son utilizados los recursos y que tan bien trabajan las unidades operativas; esto tiene lugar dentro del contexto de políticas y objetivos fijados por los planeadores estratégicos; e

involucra un intrínseco conocimiento de la toma de decisiones en la operación y consecución de tareas.

Un tercer nivel de toma de decisiones es la de control operacional, la cual involucra toma de decisiones cuando se llevan a cabo tareas específicas establecidas por planeadores estratégicos y por gerentes. Determinar cuales unidades dentro de la organización llevarán a cabo las tareas, establecer el criterio de terminación y la utilización de recursos, la evaluación de los resultados; todo esto involucra decisiones del control operacional.

### **Tipos de sistemas para decisiones**

Las dos tipos de toma de decisiones -una basada en niveles de organización y la otra en la naturaleza humana solucionadora de problemas, han sido combinadas para formar una tipología de toma de decisiones.

En general, el personal de control operacional tiene que ver con problemas generalmente bien estructurados. En contraste, los planeadores estratégicos lo tienen que hacer con problemas con escasa estructura. Aunque, cada nivel de la organización contiene tanto problemas estructurados como no estructurados.

En el pasado, la mayoría de sistemas de información exitosos tuvieron que ver con decisiones estructuradas, a niveles control operacional y gerencial.

Hoy, en la mayoría de las aplicaciones relacionadas con el control gerencial, el conocimiento y la planeación estratégica, los problemas son semiestructurados o completamente sin estructura.

### **Estructura del problema**

Es difícil encontrar problemas que sean completamente estructurados o completamente no estructurados. La vasta mayoría son semiestructurados. Esto significa que los DSS están orientados a el área donde más problemas existen.

EL DSS no intenta reemplazar al tomador de decisiones. La Figura 2.4 ilustra la relación entre la estructura del problema y el grado de apoyo que la computadora puede proveer. La computadora puede ser aplicada en la parte estructurada del problema, pero el tomador de decisiones es responsable por la parte no estructurada en la cual aplica juicio o intuición, así como análisis de conducta. El tomador de decisiones y la computadora trabajan juntos como equipo solucionador de problemas.

El objetivo de un DSS es hacer que el proceso de toma de decisiones sea lo mas eficiente posible. Cuando se busca encontrar una solución única para un problema, el proceso de toma de decisiones se apoya completamente en modelos matemáticos.

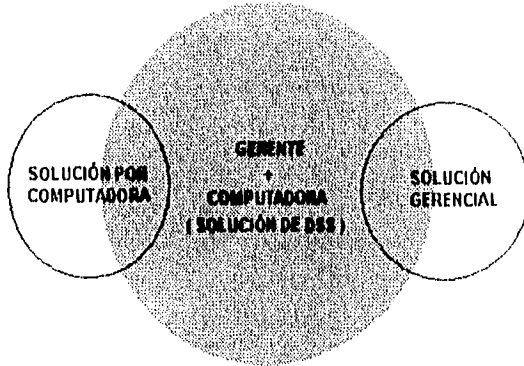


Figura 2.4 Los DSS se enfocan a los problemas semiestructurados

Cuando se buscan conocer varias posibles soluciones y con base en juicios y factores adicionales, el proceso de toma de decisiones se recaerá en aquella persona responsable de evaluar las alternativas.

#### ¿Qué son los Sistemas Expertos?

Un Sistema Experto (Expert System, ES), es un programa de computadora que utiliza el conocimiento acerca de un tópico en especial con el fin de proponer una alternativa, dar una recomendación o, si es posible una solución.

Como tal, un sistema puede satisfacer completamente la función que requiere de la experiencia humana, o bien puede jugar el rol de asistir al tomador de decisiones. El tomador de decisiones podría ser suficiente en su campo de acción, sin embargo el programa justifica su existencia desde el aspecto de apoyar a el tomador de decisiones y aumentar su productividad en disciplinas en las cuales no es un experto. Alternativamente, este elemento humano puede ser alguien quien sea capaz de alcanzar los niveles de un experto por medio de asistencia técnica recibida del sistema.

La tecnología de los sistemas expertos se deriva de la disciplina de investigación llamada Inteligencia Artificial (Artificial Intelligence, AI). Una rama de la ciencia de las computadoras referente al diseño y la implantación de programas los cuales son capaces de imitar las habilidades cognitivas, tal como lo es la solución de problemas, percepción visual y la comprensión del lenguaje. Esta tecnología ha sido exitosamente aplicada a un amplio rango de disciplinas incluyendo la química orgánica, la explotación mineral, la medicina y la estadística.

De todo lo anterior se puede obtener lo siguiente. Un sistema experto es un programa de inteligencia artificial que tiene incorporados una base de conocimiento y un sistema de inferencia. El SE es una pieza de software altamente especializado que intenta duplicar la función de un experto en algún campo determinado. El programa actúa como un consultor inteligente o consejero en un determinado dominio de interés, capturando el conocimiento de uno o más expertos. Los no expertos pueden recurrir a el sistema experto para contestar a sus preguntas, resolver sus problemas, y tomar decisiones en el dominio en que se desarrollan.

Los sistemas expertos son una nueva e innovativa forma de capturar y agrupar el conocimiento. Su fortaleza estriba en su habilidad de proporcionar información sumamente valiosa cuando el experto está ausente. Los sistemas expertos harán que el conocimiento esté más disponible y ayudará a superar el problema de traducir el conocimiento en un prácticos y útiles resultados. Y tal vez lo mejor de todo, es una más de las formas en que la tecnología ayuda a manejar la era de la información. Todo el software de IA está basado en el conocimiento, puesto que contiene hechos útiles, datos, y relaciones que serán aplicadas a el problema.

Los sistemas expertos, sin embargo, son un tipo especial de sistema basado en el conocimiento puesto que contiene conocimiento heurístico. Los heurísticos son derivados primariamente de experiencia real, no de textos. Este es conocimiento que viene directamente de cierto tipo de gente, los expertos, quienes han trabajado por determinado tiempo dentro del campo. Este es conocimiento que se deriva de el aprender haciendo. Este tal vez sea la clase más útil de conocimiento, especialmente relacionada con problemas de todos los días, que trabaja para nosotros produciendo soluciones, decisiones, y otras positivas aportaciones.

Se ha dicho que el conocimiento es poder. Ciertamente esta cita es verdadera, pero en un sentido más práctico, el conocimiento se vuelve poder solo cuando es aplicado. Los sistemas expertos son una forma más de lograr resultados más rápido y más fácil.

### **Características de un sistema experto**

Un sistema experto puede ser distinguido de la mayoría de los sistemas convencionales de aplicativos en que:

- Este simula el razonamiento humano acerca del dominio de un problema, en vez de simular el dominio por si mismo. Esto distingue a los sistemas expertos de los programas más comunes que involucran modelado matemático. Esto no quiere decir que el programa es un modelo psicológico de el experto, más bien el enfoque es imitar las habilidades del experto para solucionar problemas; esto es, ejecutar las tareas relevantes tan bien, o mejor que el experto.
- Este ejecuta razonamiento sobre representaciones de conocimiento humano, en adición a realizar cálculos numéricos o recuperación de datos. El conocimiento en el programa es generalmente expresado en algún lenguaje de propósito especial y es mantenido separado del

código que ejecuta el razonamiento. Estos módulos del programa son referidos como base de conocimiento y máquina de inferencia, respectivamente.

- Este resuelve problemas por medio de métodos de aproximación o heurística los cuales a diferencia de las soluciones algorítmicas, no garantizan el éxito. Un heurístico se compone esencialmente de reglas basadas en partes del conocimiento las cuales describen el como solucionar problemas en ciertos dominios. Estas reglas o métodos son aproximados en el sentido de que requieren de datos perfectos y que las soluciones derivadas de el sistema pueden ser propuestas con diferentes grados de certidumbre.
- Este trata con asuntos relacionados con complejidad real que normalmente requiere de un cierta cantidad de experiencia humana. Muchos programas de AI son realmente vehículos de investigación, y pueden por lo tanto enfocarse en problemas de abstracción matemática o versiones simplificadas de problemas reales para profundizar el conocimiento, o bien para depurar técnicas o métodos. Los sistemas expertos, por otro lado, resuelven problemas de genuino interés científico o comercial.
- Estos deben de exhibir alto rendimiento en términos de velocidad y confiabilidad en orden de ser una herramienta útil. Los prototipos de investigación en AI podrían no ejecutarse muy rápido, y podrían contener varias imperfecciones. Pero, un sistema experto debe de proponer soluciones en un tiempo razonable y ser correctos en la mayoría de las ocasiones, por lo menos acercarse a la frecuencia en que lo haría un experto humano.
- Este sistema debe de ser capaz de explicar y justificar soluciones o recomendaciones para convencer al usuario de que su razonamiento es correcto. Los prototipos de investigación son solo ejecutados por sus creadores o bien por personal que labora en laboratorios similares. Un sistema experto deberá poder ser utilizado por una amplia gama de usuarios, y por lo tanto debe de ser diseñado de tal forma que su ejecución sea mas bien amigable y transparente.
- El término sistema basado en el conocimiento es algunas veces utilizado como sinónimo de sistema experto aunque, estrictamente hablando, el primero es más general. Un sistema basado en el conocimiento es cualquier sistema el cual ejecuta una tarea mediante reglas adquiridas por experiencia a una representación simbólica del conocimiento, en vez de apoyarse más en métodos algorítmicos o estadísticos. Así un programa capaz de proporcionarnos información acerca de el estado del tiempo podría ser un sistema basado en el conocimiento aunque este no haya podido mostrar ningún conocimiento de meteorología, pero un sistema experto en el dominio de la meteorología debe de ser capaz de proveer pronósticos del estado del tiempo.

En suma, los sistemas expertos comprenden el conocimiento dependiente del dominio de aquellos que lo practican en un campo o tópico en particular, y usan este conocimiento para solucionar problemas, en lugar de utilizar métodos comparativamente dependientes del dominio derivados de la ciencia de las computadoras o de las matemáticas. El proceso de construir un

sistema experto es a menudo llamado ingeniería del conocimiento, y es considerada ser "la inteligencia artificial aplicada".

### **Beneficios de los sistemas expertos**

Al igual que otras formas tradicionales de software, los sistemas expertos ofrecen las siguientes ventajas :

- Permiten a los no expertos trabajar como expertos.
- Por medio de la eficiencia se mejora el trabajo y por ende, se mejora la productividad.
- Se ahorra tiempo al alcanzar los objetivos.
- Se simplifican algunas operaciones.
- Se automatizan procesos repetitivos, tediosos, y en muchas veces complejos.
- Permiten que sean resueltos nuevos tipos de problemas, esto hace que las computadoras sean útiles en nuevos campos del conocimiento.
- Capturan y almacenan conocimiento valioso que podría perderse en caso de que el experto renunciara, se jubilara o falleciera.
- Hace que el conocimiento del experto esté más difundido, lo que hace que la habilidad para resolver problemas sea adquirida por más gente.

Se puede notar que la reducción de costos no está contemplada. Los sistemas expertos, como la mayoría del software, en un principio no ahorran dinero. Se invierte en el desarrollo y uso de este, pero los beneficios derivados de este justifican el costo.

### **Desventajas**

A continuación se enuncian algunas razones las cuales demuestran que los sistemas expertos no son la panacea, que los beneficios anteriormente descritos podrían hacer creer.

- El desarrollo de un sistema expertos extremadamente difícil, más difícil que la creación de la mayoría del software convencional. No es fácil encontrar expertos. Además de que el extraerles el conocimiento a ellos no es una tarea sencilla.
- Los sistemas expertos son costosos. Es caro desarrollarlo, probarlo, y liberarlo a el usuario.
- La mayoría de los sistemas expertos de alto grado de complejidad debe de ser desarrollado en mainframes. Se pueden desarrollar sistemas expertos en computadoras personales, pero son pequeños y menos sofisticados.
- Los sistemas expertos no son 100 % confiables. Aunque los mejores expertos contribuyan a su elaboración, estos sistemas no son infalibles. Por esto mismo, su recomendación debe de ser valorada y, si es posible, probada. Una persona, tal vez, podría emitir un juicio final.

Estas desventajas son significativas, pero no imposibles de superar. Algunas de estas gradualmente irán desapareciendo en razón del desarrollo tecnológico del hardware y del software.

### **Relaciones con otro software**

Una variante es el sistema experto que se liga a un paquete de software externo como lo puede ser una hoja de cálculo, un sistema manejador de base de datos, o un programa que genera gráficos. Durante el proceso de inferencia, el sistema experto puede llegar a acceder datos en la hoja de cálculo o en el manejador de base de datos con el fin de obtener conocimiento suficiente para llegar a la solución. Un programa que genera gráficos podría ser llamado para desplegar una concepción visual de la salida obtenida. Mientras que estos programas no están realmente inmersos, el efecto es el mismo. Estos generalmente comparten memoria con el sistema experto y poseen programas que se ligan a este.

### **Tendencias**

En la actualidad la mayoría de los sistemas expertos son del tipo stand-alone. Pero en la medida de que se aplique el progreso a los sistemas expertos, existirán cada vez más diseños únicos. Todos los tipos de sistemas inmersos están proyectados a crecer considerablemente en cuanto a popularidad. De hecho, algunos predicen que el verdadero futuro de los sistemas expertos está en las aplicaciones con SE inmersos y no en los productos stand-alone.

De la misma forma que los sistemas expertos, muchas otras técnicas de AI serán empleadas como parte de un software más grande y más convencional. El lenguaje natural para los manejadores de base de datos y los paquetes de visión por computadora para sistemas robot, son ejemplos. La mejor manera de vislumbrar a la AI es como un proceso único de la ciencia de las computadoras que complementará y mejorará al software y al hardware.

### **Software de apoyo a toma de decisiones**

El software de apoyo a toma de decisiones son programas que examinarán hechos y suposiciones, las analiza, y produce una decisión o conclusión. Sin embargo, este no es un sistema experto puesto que no utiliza una representación ni una manipulación simbólica. No buscan o razonan. En vez de esto, se implantan algoritmos de toma de decisiones. Los resultados pueden ser similares a los obtenidos con un sistema experto, pero estos son alcanzados en forma diferente. Si producir una decisión correcta es gran parte del problema, sería más conveniente optar por un software de apoyo a toma de decisiones que por un sistema experto.

El software de apoyo a toma de decisiones ayuda a solucionar problemas que involucran varias alternativas. Generalmente, el usuario coloca pesos en los diferentes criterios para poder

designar sus prioridades o importancia. El software, entonces, utiliza estos pesos en un algoritmo para procesar el resultado para cada alternativa. Se ordenarán los resultados de acuerdo a probabilidad de las varias alternativas. Dados los resultados, el usuario puede tomar una decisión final.

El software de tomas de decisiones es muy útil con problemas en donde están involucradas muchas alternativas. La decisión no es muy obvia a primera vista, pero colocando factores de peso la decisión comenzará a inclinarse por determinadas alternativas.

Un software de toma de decisiones es más barato de implantar que un sistema experto, tanto en tiempo como en dinero. La desventaja podría ser que el primero requiere más atención y conocimiento para poder usarlo.

Las posibles alternativas deben de ser analizadas antes de tomar una decisión acerca de que opción utilizar.

## **SISTEMAS DE INFORMACIÓN COMO ESTRATEGIA**

En las últimas décadas, una revolución ha venido sucediendo en la manera en que organizaciones ven a la información y a los sistemas de información. Hoy en día, compañías y organizaciones están utilizando la tecnología de información como una arma competitiva para desarrollar nuevos productos y servicios, forjan nuevas relaciones con los proveedores y competidores, cambiar radicalmente sus operaciones así como sus estructuras internas. Al paso del tiempo, los gerentes encargados de funciones dedicadas al negocio se orientarán hacia las computadoras, los gerentes de sistemas de información se orientarán mas hacia los negocios y, el problema de utilizar el mismo lenguaje será cosa del pasado.

Una alternativa viable para afrontar los problemas inherentes de los sistemas de información es el utilizar la planeación estratégica.

Durante la transición, hacia la planeación estratégica, se puede crear un grupo de apoyo tecnológico. Este grupo estará al pendiente, no solo las tecnologías emergentes para su posible uso dentro de la organización sino que también de aquellas tecnologías que se estén utilizando actualmente. El grupo de apoyo tecnológico puede consistir de personas con experiencia en el negocio, así como un razonable conocimiento de computación. Su objetivo podría ser explorar las formas para utilizar la tecnología de información, y así esta se convierta en un factor de ventaja.

### **¿Qué significa estratégico?**

Se les llama sistemas estratégicos, a aquellos sistemas que modifican los objetivos, productos, servicios o el ambiente de las organizaciones. Los sistemas que cumplen estas características literalmente cambian el negocio y su organización.



## **Rol de los sistemas de información en la estrategia**

En muchas organizaciones la alta dirección ha visualizado a los sistemas de información sólo como una actividad de soporte. Esta operación ha venido siendo importante para la compañía, pero al parecer está teniendo poco o nulo impacto en áreas principales para la toma de decisiones. Sin embargo, este punto de vista está cambiando en cada vez más y más compañías. La competencia por la permanencia, o bien por la supremacía en cada mercado está llevando a los sistemas de información a ser un foco de atención. Ellos pueden influir y están influyendo en los ingresos, los nuevos productos, los nuevos mercados, y las adquisiciones en general. Entonces, estos están teniendo impacto en la competitividad de la organización. En suma, los sistemas de información pueden ser utilizados para que en conjunción con los demás elementos de la organización, realicen un esfuerzo de cooperación y así mejorar la competitividad de la misma. Esta forma de trabajar es relativamente nueva, sobre todo como opción competitiva.

Los sistemas de información también pueden tener un mayor impacto en los procedimientos internos de la organización. Tales impactos pueden cambiar significativamente la forma en que las compañías operan y, en que tan efectivas son.

### **¿Qué son los sistemas estratégicos?**

Las definiciones de sistemas estratégicos abundan. Se puede definir un sistema estratégico como un sistema que apoya o bien da forma a unidades de negocio dedicadas a la estrategia. Los sistemas estratégicos, poseen un aspecto externo, el cual está dirigido a proveer nuevos servicios a los clientes y/o proveedores. Los sistemas estratégicos poseen, también, un aspecto interior, el cual está dirigido a mejorar la posición competitiva de la firma mediante el incremento de la productividad de los empleados. Los sistemas electrónicos de oficina son generalmente citados como este tipo de sistemas estratégicos.

Un sistema estratégico es aquel que posee las siguientes tres características.

1. Un sistema estratégico cambia significativamente el desempeño del negocio.
2. Un sistema estratégico colabora con la obtención de una meta estratégica.
3. Un sistema es estratégico, si este fundamentalmente cambia la manera en que la compañía hace negocios, la manera en que compete, o la manera en que la compañía trata con sus clientes o sus proveedores.

Existen por lo menos dos diferencias entre los sistemas tradicionales y los sistemas estratégicos. Una de estas diferencias es que las técnicas para el manejo del proyecto tradicional no trabajan bien para desarrollar sistemas estratégicos. La otra diferencia es que el involucramiento de el usuario o cliente es de vital importancia puesto que el sistema puede llegar a cambiar la manera en que el negocio opera.

## **Manejo estratégico**

Manejo estratégico es el proceso de determinar la futura dirección de una organización y de implementar políticas que guiarán a la firma en esa dirección .

Sin una adecuada planeación estratégica, todos los días decisiones operativas serán tomadas sin referencia a algún lineamiento general. En esta situación, aunque las decisiones parezcan estar en un contexto correcto pueden entrar en conflicto con decisiones tomadas en otra área. Un plan estratégico beneficia porque provee un ambiente predecible en el cual tomar decisiones.

El proceso de planeación estratégica es similar al establecer una casa. Una mejor y más sólida base redundará en una estructura semejante. La planeación estratégica se puede dar en varios niveles de una organización. El mas alto de estos es el nivel corporativo, este nivel de planeación será responsabilidad de la alta dirección. Las metas y objetivos a este nivel involucran a la dirección de la compañía por completo. Estos planes, también, involucran mayores compromisos de recursos y tendrán un efecto sustancial en la organización.

El siguiente nivel de planeación estratégica será llamado nivel de negocios, y el nivel final será llamado nivel funcional. El nivel de negocio involucra la planeación de líneas específicas de negocios. Este podría incluir producción, comercialización, distribución, etc. Los planes a este nivel deben ser congruentes, con los planes corporativos y los planes para su unidad específica de negocio.

Este proceso de planeación será tan interactivo como lo sea el grado de involucramiento de la alta dirección con la de nivel intermedio, en el proceso de decisiones. De esta forma ideas innovativas pueden provenir de mas individuos. De la misma manera, habrá mas aceptación de los planes que los involucran y afectan directamente. En adición, el proceso de planeación es un proceso iterativo. Los planes serán desarrollados inicialmente y después analizados y revisados de acuerdo a como tengan que ser adecuados.

## **Planeación de sistemas de información**

Para que cualquier organización funcione de una manera racional, es necesario tener un plan de acción a largo plazo. Para la alta dirección, este plan a largo plazo es especificado por una serie de metas organizacionales. El cumplimiento de muchas de estas depende de la productividad de los sistemas de cómputo. De aquí, que es crítico que cualquier plan de negocios incluya a el área de sistemas de información.

El plan para sistemas de información debe de ser general. La esencia de cualquier plan a largo plazo es dar a los individuos una amplia idea de la futura dirección de la organización sin restringir excesivamente la capacidad de toma de decisiones de aquellos que manejan la organización. Mientras un plan nos puede describir anticipadamente las áreas de crecimiento, las

de contracción así como la magnitud de estas, este no puede forzar a que las decisiones sean las apropiadas y en el tiempo adecuado.

La planeación estratégica es una actividad consumidora de tiempo. Esto y el hecho de que muchos gerentes resienten la carga del manejo de la operación del día a día, hacen difícil la decisión de comenzar a desarrollar un plan estratégico. Sin objetivos a largo plazo, las decisiones serán hechas solo considerando consecuencias a corto plazo.

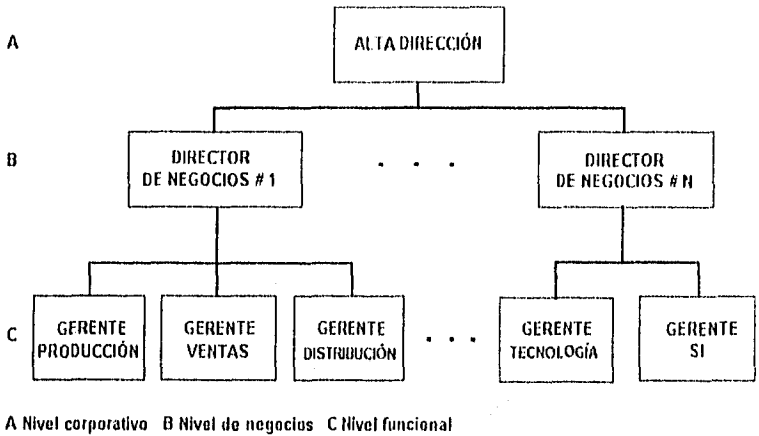


Figura 2.5 Niveles de manejo estratégico

### Costos y beneficios de la planeación

Como en cualquier decisión de negocio, existen costos y beneficios asociados con la planeación estratégica. Para la mayoría de las firmas los beneficios de la planeación estratégica sobrepasan a los posibles costos.

Los siguientes son algunos beneficios de la planeación estratégica.

Un plan estratégico establece un ambiente más predecible en el cual las decisiones pueden ser tomadas. El proceso de planeación permite que los problemas puedan ser analizados y que soluciones alternativas sean investigadas antes de que se conviertan en algo crítico. Si se establecen metas los empleados podrán estar más motivados y así promover la cooperación entre varias áreas. Las metas también proveen de herramientas más objetivas y útiles, en la evaluación y control dentro de la organización. Finalmente, la planeación estratégica es obviamente la opción lógica a seguir.

Por el lado de los costos, el proceso de planeación puede ser un consumidor de tiempo. Esto

puede tener un efecto adverso en las operaciones de la organización si estos procesos no tienen un estimado de tiempo adecuado. Si el plan estratégico resultante no es seguido, entonces el proceso de planeación será visto como un procedimiento burocrático que consume tiempo valioso. El proceso de planeación puede también convertirse en algo político. Por supuesto la política podría llegar a ser un asunto asociado con todos los aspectos del negocio y, tal vez llegar a causar algún inconveniente. El equipo de planeación debe estar alerta para identificar a este tipo de problemas.

### **El proceso de planeación**

Dentro de cada unidad organizacional la planeación puede ser realizada de diferentes maneras. Se puede crear un equipo cuya única función sea la responsabilidad de la planeación estratégica. Este equipo podría reportarle directamente al gerente de la unidad. Alternativamente, la planeación puede ser hecha por un comité compuesto de ciertos gerentes de unidad.

El proceso típico de planeación incluye varios pasos. El primer paso es definir tanto en ambiente interno como el externo. El ambiente externo puede ser, a su vez, separado en ambiente directo e indirecto. Como su nombre lo dice, el ambiente externo directo incluye todos los factores que están fuera de la unidad de planeación y que afectarán directamente a esta. La definición de estos ambientes depende de el nivel al cual la planeación sea hecha. El ambiente interno incluye todos los recursos y actividades dentro de la unidad organizacional.

Es necesario analizar los ambientes interno y externo antes de que sea formulado cualquier plan estratégico. De otra manera, esto será como tratar de realizar una travesía sin conocer el punto de partida y su entorno. El proceso de analizar estos ambientes es llamado revisión ambiental.

Dentro del contexto de ambiente, la misión de cada unidad debe de ser determinada. La misión es una expresión de amplios límites. Esto debe de incluir los tipos de productos y servicios ofrecidos.

De aquí que el proceso se comience a convertir más específico y detallado. La idea es desarrollar un criterio, con el cual los gerentes de nivel medio se basen para operar. El siguiente paso es seguir de la manera más fiel lo antes establecido. Este involucrará el desarrollo de estrategias dentro de los límites de la misión establecida. De estas alternativas las metas estratégicas pueden ser determinadas. Las metas estratégicas son propuestas generales y por naturaleza no son cuantitativas. Estas metas son creadas de manera más exacta mediante la creación de objetivos estratégicos, y estos objetivos estratégicos son vueltos operativos por los objetivos de la compañía basados en ciclos de tiempo.

Aunado con el proceso de planeación está el desarrollo de políticas organizacionales y de estándares. A diferencia de los objetivos, las políticas y los estándares no son cuantitativas. En vez

de esto, ellos dictan las maneras en que ciertas situaciones deben de ser manejadas. Como muestra la Figura 2.6, las políticas y los estándares pueden ser directamente relacionadas con la implantación de objetivos estratégicos o metas estratégicas. Y, además estrechamente ligado con la implantación de metas y/o objetivos estratégicos esta la estructura de la organización.

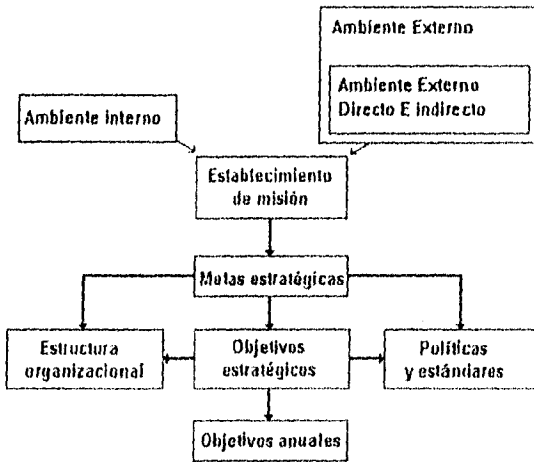


Figura 2.6 Proceso de planeación estratégica

De muchas maneras la estructura organizacional y el grado de establecimiento de políticas y estándares están íntimamente relacionadas. Una estructura administrativa más centralizada estará asociada con más políticas y estándares.

Las organizaciones descentralizadas permitirán más independencia para las unidades organizacionales más bajas. Si es hecho propiamente, el proceso de planeación estratégica debe de ser extenso, e incluir a toda la compañía. En suma, las metas estratégicas deben de ser periódicamente revisadas para que puedan ser realizados ciertos ajustes cuando las condiciones internas y/o externas se modifiquen.

### Metas y objetivos

Para cualquier conjunto de planes que vayan a ser utilizados, deben de existir un conjunto de metas y objetivos con base en los cuales los planes serán fijados.

Una meta generalmente implica un resultado hacia el cual las actividades son direccionadas mientras que un objetivo es un evento contra el cual es medida la efectividad de la metas. No es

suficiente establecer una meta sin que se establezca, también, algún mecanismo que pueda ser utilizado para medir los alcances de esta meta.

Un centro de proceso de datos debe de establecer metas y objetivos a corto y a largo plazo. Este proceso no es completamente autónomo, para poder brindar un servicio efectivo a la organización, estas metas y objetivos deben de encajar bien en la naturaleza de la compañía. Una mejor compaginación, implicará por fuerza una mayor probabilidad de éxito.

Para las metas acogidas por el área de sistemas de información, debe de existir por fuerza una comunicación efectiva con los niveles superiores. Los objetivos deben de ser realistas.

Todas las partes involucradas deben de participar activamente. La comunicación es esencial. Los objetivos deben de ser claros. Deben de existir estándares. El apoyo de los niveles superiores medios, bajos y por toda la organización es crucial. Se debe de hacer una elección inteligente de las metas y objetivos. Los criterios para establecer las metas y los objetivos nunca deberán de ser estáticos. Una vez que han sido fijados estos deben de ser entendidos y difundidos por toda la organización.

### **Tecnología de información como agente de cambio organizacional**

Los sistemas son diseñados para servir en varios aspectos a una organización. En su papel de apoyo, sostienen a la organización existente, los productos existentes, y los procedimientos de negocio existentes.

Los sistemas de información continúan protagonizando este papel, pero además se esta incrementando el uso de estos para lograr el cambio de la organización incluyendo sus productos y procedimientos. En este nuevo rol, la tecnología y los sistemas guían las organizaciones hacia nuevos patrones de conducta. Las organizaciones deben de cambiar su organización interna para aprovechar las ventajas de la nueva tecnología de información. Utilizando los sistemas de información estratégicamente puede requerir cambios drásticos en la organización, como lo pueden ser nuevos gerentes, una nueva fuerza de trabajo y una relación más estrecha con clientes y proveedores. □

### ¿QUÉ ES Y PARA QUÉ SIRVE LA INFERENCIA ESTADÍSTICA?

□ En estadística se le llama inferencia al proceso de inducir o bien deducir las características o parámetros poblacionales a partir de información muestral, midiendo con probabilidades la incertidumbre inherente.

La teoría de la inferencia estadística consiste en aquellos métodos con los cuales se pueden realizar inferencias acerca de una población. La tendencia para estimar un parámetro poblacional es distinguir entre el método clásico y el método bayesiano. En el método clásico las inferencias se basan en la información obtenida de una muestra aleatoria seleccionada de la población. El método bayesiano utiliza un conocimiento subjetivo previo acerca de la distribución de la probabilidad con parámetros desconocidos, junto con información proporcionada por los datos muestrales.

Si no existiera la inferencia estadística, no sería posible diseñar reglas confiables para un control de calidad. Por ejemplo, prácticamente sería imposible determinar si todas las varillas que serán utilizadas en una construcción satisfacen las especificaciones de diseño, y mucho menos, si para probarlas se somete a pruebas destructivas.

Para hacer una inferencia estadística, se pueden estimar los parámetros descriptivos de la población a partir de la información obtenida en una muestra, ya sea como un valor puntual, como un intervalo, o bien establecer valores hipotéticos de los parámetros y probar estadísticamente si son válidas estas hipótesis.

A los dos primeros casos se les llama estimación estadística puntual y por intervalos, y al último se le llama prueba de hipótesis.

De aquí que la inferencia estadística puede dividirse en dos áreas principales: estimación y pruebas de hipótesis. Para distinguir claramente entre ambas, considérense los siguientes ejemplos.

Un candidato para un puesto público desea estimar la proporción real de votantes que lo apoyan, mediante la obtención de las opiniones de una muestra aleatoria de cien votantes. La fracción de ellos que lo apoye puede utilizarse como una estimación de la proporción real de la población total de votantes. Un conocimiento de la distribución muestral de una proporción permite establecer el grado de precisión de la estimación. Este problema pertenece, entonces, al área de estimación.

Ahora considérese el caso en el cual una ama de casa se interesa en determinar si el detergente para ropa de la marca Ariel es más resistente a la suciedad que el detergente de la marca Tide. Esta ama de casa podría suponer que la marca Ariel es mejor que Tide y, después de realizar las pruebas apropiadas, aceptar o rechazar esa hipótesis. En este ejemplo no se intenta estimar un parámetro, sino tomar una decisión correcta respecto a la hipótesis preestablecida. Depende de la teoría del muestreo para obtener alguna medida de precisión para la decisión que se tome.

En cualquier actividad profesional, se presenta frecuentemente la necesidad de describir el comportamiento de fenómenos aleatorios, tales como la resistencia de un material, la respuesta de cierto sistema físico o la demanda de un producto. Para lograrlo se puede partir de la información obtenida de una muestra y aplicando la inferencia estadística.

Antes de comenzar con la inferencia estadística es conveniente apuntar algo sobre modelos probabilísticos.

## MODELOS PROBABILÍSTICOS

Cuando se estudia un experimento aleatorio mediante la teoría de la probabilidad y la estadística, es necesario definir el comportamiento de las variables involucradas. Una forma de hacerlo es mediante su función masa o densidad de probabilidad; sin embargo, encontrar para cada variable su verdadera distribución de probabilidad es una tarea muy difícil y por tal motivo, se han desarrollado una gran cantidad de modelos probabilísticos que explican satisfactoriamente el comportamiento de muchas variables aleatorias.

A continuación se presentan algunos de los modelos teóricos, tanto para variables aleatorias discretas como continuas, incluyendo su forma, sus parámetros y otras características particulares.

## DISTRIBUCIONES DISCRETAS

Algunos de los modelos que se presentan en este capítulo se han desarrollado a partir de un experimento denominado ensayo de Bernoulli, en el cual sólo se pueden presentar dos posibles resultados, denominados éxito y fracaso. La probabilidad de que ocurra un fracaso con  $q$  ó  $1 - q$ .



Como ejemplos de un ensayo de Bernoulli se puede considerar el lanzamiento de una moneda, cuyos posibles resultados son águila o sol, o bien tomar un producto de una línea de producción para verificar si cumple con las normas de calidad preestablecidas, en cuyo caso los posibles resultados son bueno o defectuoso.

Llevarle éxito a uno de los posibles resultados no significa forzosamente que sea el más favorable, generalmente se le llama así al evento que tiene menos probabilidad de ocurrencia; por ejemplo, se puede llamar éxito al posible evento de que un vehículo se descomponga durante un intervalo de tiempo y fracaso, si no sufre ningún daño.

### Proceso o secuencia de Bernoulli

Consiste en una serie de experimentos o ensayos de Bernoulli independientes, donde la probabilidad de éxito  $p$  permanece constante durante todo el proceso. Por las características anteriores, se dice que el proceso es dicotómico, estacionario e independiente.

### Distribución de Bernoulli

Si  $X$  es el número de éxitos en un ensayo de Bernoulli con probabilidad  $p$  de éxito y  $1-p$  ó  $q$  de fracaso, entonces se dice que  $X$  tiene distribución binaria o de Bernoulli con parámetro  $p$ .

Como  $X$  solamente puede tomar el valor de cero o uno, su función masa de probabilidad está definida por:

$$P_X(x) = \begin{cases} 1-p & \text{para } x = 0 \\ p & \text{para } x = 1 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

La media o valor esperado de esta distribución, que se puede obtener de acuerdo con las propiedades de la función generatriz de momentos, es:

$$\mu_x = p$$

la variancia sería:

$$\sigma_x^2 = pq$$

Aun cuando esta distribución no se aplica por sí sola en el estudio de fenómenos de la realidad, es la parte fundamental de los modelos basados en procesos bernoullianos.

Conviene destacar el hecho de que ésta y las demás distribuciones que se presentan en este capítulo son en realidad familias de funciones, debido a que para cada valor del parámetro (o parámetros) se obtiene diferentes funciones que pertenecen a la misma familia.

### Distribución binomial

Sea  $k$  una variable aleatoria que representa el número de éxitos en una secuencia de  $n$  ensayos de Bernoulli con probabilidad  $p$  de éxito y  $1-p$  de fracaso, entonces  $k$  tiene distribución

binomial con parámetros  $n$  y  $p$ .

Para determinar la función masa de probabilidad de la distribución binomial, supóngase que se desea determinar la probabilidad de encontrar  $k$  piezas defectuosas de un lote de tres unidades, tomadas de una de las líneas de producción de una planta industrial. Considérese que la probabilidad de que una pieza resulte defectuosa es  $p$ , esto es  $p$  (éxito) =  $p$  y  $p$  (fracaso) =  $1 - p$ .

El número de maneras en las cuales se pueden presentar  $k$  éxitos en una serie de  $n$  ensayos, es el número combinatorio:

$$C(n, k) = \frac{n!}{k!(n-k)!}$$

y la probabilidad de cada uno de los posibles resultados es  $p^k(1-p)^{n-k}$  por lo cual la F.M.P. de la distribución binomial se representa en forma general como:

$$P_X(k) = \begin{cases} C(n, k) p^k q^{n-k} & \text{para } k \leq 0 \\ 0 & k < 0 \end{cases}$$

donde  $k$  y  $n$  son enteros positivos y las constantes  $n$  y  $p$  son los parámetros de esta distribución.

Para obtener la media y la variancia de la distribución binomial, nuevamente conviene obtener su función generatriz de momentos, si se desarrolla matemáticamente se obtendrán, la media y la variancia:

$$\mu_x = np \qquad \sigma_x^2 = npq$$

### Distribución Pascal

Sea  $n$  una variable aleatoria que representa el número de ensayos necesarios para encontrar  $k$  éxitos en una secuencia de Bernoulli con probabilidad de éxito  $p$ , entonces  $n$  tiene distribución binomial negativa con parámetros  $k$  y  $p$ .

Comparando esta definición con la correspondiente a la binomial, se debe notar que el número de éxitos  $k$  es una variable en la distribución binomial, mientras que en esta distribución es solamente un parámetro. También se debe observar que en esta distribución, el último ensayo de la secuencia tiene que ser un éxito para completar los  $k$  éxitos que aparecen en la definición.

Tomando en cuenta esta última observación, se puede obtener la F.M.P.

$$P_N(n) = \begin{cases} C(n-1, k-1) p^k q^{n-k} & \text{para } n \geq k \\ 0 & \text{para } n < k \end{cases}$$

Siguiendo un procedimiento semejante al practicado en la distribución binomial, se obtiene de la función generatriz de momentos se obtienen la media y la variancia:

$$\mu_n = \frac{k}{p} \qquad \sigma_n^2 = \frac{kq}{p^2}$$

### Distribución de Poisson

Si se divide el intervalo de definición de una variable continua  $t$  que puede representar tiempo, distancia o cualquier otra variable continua en pequeños subintervalos iguales  $\Delta t$ , si en cualquier punto  $t$  se puede presentar dos posibles resultados denominados éxito o fracaso, entonces se puede considerar a cada subintervalo  $\Delta t$  como probabilidad  $p$  (en cualquier  $\Delta t$ ) o un fracaso con probabilidad  $q = 1 - p$ .

Cuando la ocurrencia de un éxito en un subintervalo es independiente de los que ocurre previamente y no influye en lo que ocurre posteriormente, entonces el intervalo de definición de  $t$  representa un proceso de Bernoulli.

Si además, el tamaño de los subintervalos  $\Delta t$  tienden a cero y la tasa promedio de éxitos por unidad de  $t$  definida como  $\lambda = n p$  permanece constante, se trata entonces de un proceso Poisson.

Algunos ejemplos de dicho proceso se presentan al considerar como variable continua el tiempo y como éxito la llegada de un cliente a un almacén, la descompostura de uno de los elementos de cierto mecanismo o la detección de un sismo.

Si una variable aleatoria  $Y$  representa el número de éxitos en un proceso Poisson con una tasa promedio de éxitos  $\lambda$ , entonces  $Y$  tiene distribución de Poisson con parámetro  $\lambda$ .

Cuya función masa de probabilidad es:

$$P_Y(y) = \begin{cases} \frac{\lambda^y e^{-\lambda}}{y!} & \text{para } y \geq 0 \\ 0 & \text{para } y < 0 \end{cases}$$

Siguiendo un procedimiento similar al de las distribuciones anteriores, de la función generatriz de momentos, se obtienen:

$$\mu_Y = \lambda \qquad \sigma_Y^2 = \lambda$$

### DISTRIBUCIONES CONTINUAS

Hasta ahora se han presentado modelos probabilísticos que se basan en los procesos de Bernoulli o Poisson y manejan variables aleatorias discretas. En esta sección se presentarán modelos probabilísticos para variables aleatorias continuas, que no se basan necesariamente en los procesos antes mencionados.

#### Distribución uniforme o rectangular

Sea  $X$  una variable aleatoria que puede tomar, con la misma probabilidad, cualquier valor en el intervalo  $[a, b]$ , entonces  $X$  tiene distribución uniforme con parámetros  $a$  y  $b$ .

En la realidad existen muy pocos experimentos cuyos resultados tengan la misma probabilidad de ocurrencia; sin embargo, en algunos casos se utiliza esta distribución cuando no se cuenta con la información necesaria y se tienen que asignar probabilidades a los eventos. Por ejemplo, la proporción de personas que consumirán un nuevo producto en una comunidad, se puede representar mediante una variable aleatoria con distribución uniforme y parámetros  $a = 0$  y  $b = 1$ .

Para la esta distribución se tiene que:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Obteniendo los momentos de la distribución uniforme estos son, para la media y la variancia:

$$\mu_X = \frac{1}{2}(b+a) \quad \sigma_X^2 = \frac{(b-a)^2}{12}$$

### Distribución normal y normal estándar

La distribución normal o de Gauss es uno de los modelos probabilísticos de más utilidad debido entre otras razones, a que en la realidad muchas variables aleatorias pueden representarse adecuadamente mediante esta distribución, incluso otras distribuciones que se estudiarán en este capítulo toman como base a la distribución normal y además, bajo ciertas condiciones, la distribución normal se aproxima a otras distribuciones teóricas.

En forma general se dice que si la función densidad de probabilidad de una variable aleatoria  $X$  está dada por:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

entonces  $X$  tiene distribución normal, con parámetros  $\mu$  y  $\sigma$ .

La gráfica por su forma es llamada campana de Gauss cuya función es simétrica con respecto a la media; su punto máximo está en  $X = \mu$  y es asíntotica con el eje  $X$  (en ambos lados).

Conforme a la definición, de la función generatriz de esta distribución obtenemos media y variancia:

$$\mu_X = \mu \quad \sigma_X^2 = \sigma^2$$

Es decir, que la media y la variancia de la distribución normal son respectivamente los parámetros  $\mu$  y  $\sigma^2$ .

Para calcular la probabilidad de que la variable aleatoria  $X$  tome un valor dentro del intervalo

$[x_1, x_2]$  se debe evaluar la integral definida:

$$P(x_1 \leq X \leq x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

la cual presenta algunos problemas para obtenerla analíticamente.

Haciendo el cambio de variables (estandarizando la variable)

$$Z = \frac{x - \mu_x}{\sigma_x}$$

Se puede demostrar que  $Z$  también tiene distribución normal con parámetros  $\mu_z = 0$  y  $\sigma_z = 1$

A la distribución de probabilidad de la variable  $z$  se le denomina normal estándar y su función de distribución acumulada es:

$$F_z(z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{-\frac{z^2}{2}} dz$$

Para calcular la probabilidad de que  $z$  tome un valor en cierto intervalo, se puede aprovechar la simetría de la función y el hecho de que el área bajo la curva que se encuentra a uno y otro lado del eje de simetría es 0.5.

### Distribución exponencial

Sea  $T$  una variable aleatoria que representa el tiempo que transcurre entre la ocurrencia de dos éxitos consecutivos de un proceso Poisson, con una tasa promedio de éxitos  $\lambda$ ; entonces  $T$  tiene distribución exponencial con parámetro  $\lambda$ .

La función densidad de probabilidad de la distribución exponencial es:

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t \geq 0 \\ 0 & \text{para } t < 0 \end{cases}$$

De la función generatriz de momentos para la media se tiene:

$$\mu_T = \frac{1}{\lambda} \quad \sigma_T^2 = \frac{1}{\lambda^2}$$

## INFERENCIA ESTADÍSTICA

### Estimación de parámetros

La estimación de un parámetro involucra el uso de datos muestrales en conjunción con algún estadístico. Existen dos formas de llevar a cabo lo anterior; la estimación puntual y la estimación por intervalo. En la primera se busca un estimador que, con base en los datos muestrales, dé origen a una estimación univalueada. Es decir, cuando solamente se da un valor del parámetro

conocido; este valor numérico se calcula a partir de una muestra dada y sirve como una aproximación del valor exacto del parámetro, el cual es desconocido. Para la segunda, se determina un intervalo en el que, de forma probable, se encuentre el valor del parámetro. Este intervalo recibe el nombre de intervalo de confianza.

### Estimadores puntuales

Las medidas descriptivas de una muestra se usan generalmente como estimadores puntuales de los parámetros poblacionales; sin embargo, otro aspecto importante se refiere a la estimación de parámetros como  $p$  cuando se sabe que la población es binomial, o  $\lambda$  si tiene distribución de Poisson.

En esta sección se definen algunas características de los estimadores puntuales que se utilizan para seleccionar, de entre varios de ellos, el más conveniente. También se presentan métodos para obtenerlos.

Entre las características más importantes de un estimador se encuentran el sesgo, la eficiencia y la consistencia. A continuación se definen cada una de ellas.

### Propiedades deseables de los estimadores puntuales

El estadístico que se utiliza para obtener una estimación puntual recibe el nombre de estimador o función de decisión. De aquí que la función de decisión  $S^2$ , la cual es una función de la muestra aleatoria, es un estimador de  $\sigma^2$  y la estimación  $s^2$  es la acción que se toma. Generalmente muestras diferentes conducen a acciones o estimaciones diferentes.

No se espera que un estimador obtenga sin error, del parámetro poblacional. Tampoco que  $\bar{X}$  estime el valor de  $\mu$  con exactitud, sino que no se aleje mucho del valor real. Para una muestra en particular es posible tener una estimación más precisa de  $\mu$  utilizando la mediana muestral como estimador. Considérese, por ejemplo, una muestra que comprenda los valores 2, 5, 11 de una población cuya media es 4 aunque supuestamente se le desconoce. Se podría estimar que  $\mu$  fuera  $\bar{x} = 6$ , utilizando la media muestral como estimación, o  $\bar{x} = 5$ , utilizando la mediana muestral como tal. En este caso el estimador  $\bar{X}$  produce una estimación más cercana al parámetro real que la del estimador  $\bar{X}$ . Por otra parte, si la muestra aleatoria contiene los valores 2, 6 y 7, entonces  $\bar{x} = 5$  y  $\bar{x} = 6$ , de tal manera que  $\bar{X}$  es ahora el mejor estimador. Sin conocer el valor real de  $\mu$ , se debe decidir de antemano si se empleará  $\bar{X}$  o  $\bar{X}$  como estimador.

Es posible definir muchos estadísticos para estimar un parámetro desconocido. Entonces, ¿Cómo seleccionar un buen estimador? ¿Cuáles son los criterios para juzgar que un estimador es bueno o malo? De manera intuitiva, ¿Qué es un buen estimador? Si se piensa en términos de estimadores humanos como los que se encuentran en las grandes compañías de construcción,

entonces quizá un buen estimador sea aquella persona cuyas estimaciones siempre se encuentran muy cercanas a la realidad.

Si se acepta la premisa de que el objetivo de la estimación de parámetros no es igual al de los estimadores o predictores humanos. humano. Del análisis, surgen dos propiedades deseables: el estimador de un parámetro  $\theta$  debe tener una distribución de muestreo concentrada alrededor de  $\theta$  y la variancia del estimador debe de ser la menor posible.

### Estimadores insesgados

¿Cuáles son las propiedades de una "buena" función decisión que influirían en la selección de un estimador sobre otro? Sea  $\hat{\theta}$  un estimador cuyo valor  $\hat{\theta}$  es una estimación puntual de algún parámetro poblacional desconocido  $\theta$ . En verdad sería deseable que la distribución muestral de  $\hat{\theta}$  tuviera una media igual al parámetro estimado. Se dice que un estimador que posee esta propiedad es insesgado.

Se dice que un estadístico  $\hat{\theta}$  es un estimador insesgado del parámetro  $\theta$  si,  $E[\hat{\theta}] = \theta$

### Estimador eficiente

Si  $\hat{\theta}_1$  y  $\hat{\theta}_2$  son dos estimadores insesgados del mismo parámetro poblacional  $\theta$ , se seleccionaría el estimador cuya distribución muestral tuviera la variancia más pequeña. De aquí que si  $\sigma_{\hat{\theta}_1}^2 < \sigma_{\hat{\theta}_2}^2$ , se afirma que  $\hat{\theta}_1$  es un estimador más eficiente de  $\theta$  que  $\hat{\theta}_2$ . Si se consideran todos los estimadores insesgados posibles de algún parámetro  $\theta$ , aquel con la variancia más pequeña recibe el nombre de estimador más eficiente de  $\theta$ .

### Estimador consistente

Es razonable esperar que un buen estimador de un parámetro  $\theta$  sea cada vez mejor conforme crece el tamaño de la muestra. Esto es, conforme la información en una muestra aleatoria se vuelve más completa, la distribución de muestreo de un buen estimador se encuentra cada vez más concentrada alrededor del parámetro  $\theta$ . Se tendrá un mejor estimador de  $\theta$  si se basa en treinta observaciones que si lo hace con sólo cinco. Esta idea origina lo que se conoce como un estimador consistente.

Un estimador  $\hat{\theta}_n$  obtenido de una muestra de tamaño  $n$  es consistente si  $\hat{\theta}_n$  converge en probabilidad a  $\theta$ , es decir:

$$\lim_{n \rightarrow \infty} P \{ |\hat{\theta}_n - \theta| > \epsilon \} = 0$$

para cualquier constante positiva  $\epsilon$ .

### Intervalos de confianza

En algunos casos no es suficiente estimar un parámetro de manera puntual, debido a que el estimador es una variable aleatoria y la probabilidad de que tome el mismo valor que el parámetro es prácticamente cero.

Para aumentar la probabilidad de que la estimación coincida con el parámetro deseado, se debe de establecer un intervalo denominado intervalo de confianza que contenga, con cierta probabilidad, al parámetro deseado.

A los límites del intervalo de confianza se les denomina límites de confianza y a la probabilidad de que el parámetro de la población esté en el intervalo de confianza se le llama nivel de confianza de la estimación y, en lo sucesivo, se representará como  $(1 - \alpha)$ .

Para establecer los límites de confianza (LC) de un parámetro  $\theta$ , se parte de una distribución muestral. En la figura 3.1 se pueden observar los elementos que intervienen en la estimación muestral.

Se puede notar que el complemento del nivel de confianza  $\alpha$  se distribuye en dos partes iguales a los lados del intervalo, con el objeto de centrarlo.

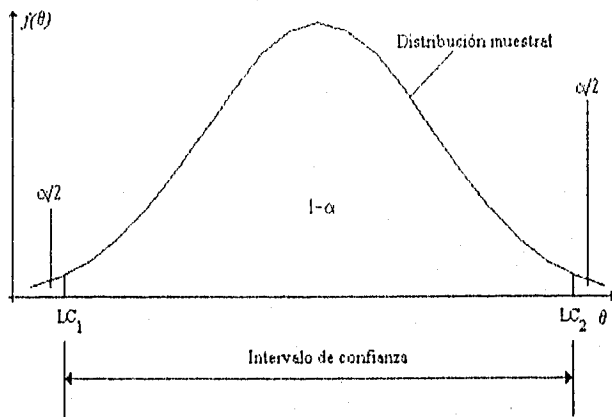


Figura 3.1 Regiones críticas y de aceptación de prueba de hipótesis para media de población normal

En términos generales, la construcción de un intervalo de confianza para un parámetro desconocido  $\theta$  consiste en encontrar un estadístico suficiente  $T$  y relacionarla con otra variable aleatoria  $X = f(T; \theta)$ , en donde  $X$  involucra a  $\theta$  pero la distribución de  $X$  no contiene a  $\theta$ , así como tampoco a ningún otro parámetro desconocido. Entonces se seleccionan dos valores de  $x_1$  y  $x_2$  tales que

$$P(x_1 < X < x_2) = 1 - \alpha$$



en donde, como ya se vio,  $1 - \alpha$  recibe en nombre de coeficiente de confianza. Mediante una manipulación algebraica de las dos expresiones, se puede modificar el contenido entre paréntesis y expresarlos como

$$P[h_1(T) < \theta < h_2(T)] = 1 - \alpha$$

en donde  $h_1(T)$  y  $h_2(T)$  son funciones del estadístico  $T$  y de esta forma, variables aleatorias. El intervalo de confianza para  $\theta$  se obtiene sustituyendo en  $h_1(T)$  y  $h_2(T)$  los estimadores calculados a partir de los datos muestrales, dando origen a lo que se conoce como intervalo de confianza bilateral. Al seguirse el mismo procedimiento, también pueden desarrollarse intervalos de confianza unilaterales, de la forma

$$P[g_1(T) < \theta] = 1 - \alpha$$

o

$$P[\theta < g_2(T)] = 1 - \alpha$$

El primero es un intervalo de confianza unilateral inferior para  $\theta$ , y el segundo es un intervalo de confianza unilateral superior.

## PRUEBA DE HIPÓTESIS ESTADÍSTICA

Hasta este punto ya se ha examinado la inferencia estadística con respecto a la estimación puntual y por intervalo. Ahora se revisará la otra parte de la inferencia: la prueba o contraste de una hipótesis estadística. Como se verá, la prueba de una hipótesis estadística tiene una fuerte relación con el concepto de estimación.

Una hipótesis estadística es una afirmación con respecto a alguna característica desconocida de una población de interés. La esencia de probar una hipótesis estadística es el decidir si la afirmación está apoyada por la evidencia experimental, la cual se obtiene a través de la muestra aleatoria. En forma general, la afirmación involucra ya sea a algún parámetro o a alguna forma funcional no conocida de la distribución de interés a partir de la cual se obtiene una muestra aleatoria. La decisión acerca de que si los datos muestrales apoyan estadísticamente la afirmación se toma con base en la probabilidad, y, si esta es mínima, entonces es rechazada.

### Hipótesis estadística

De acuerdo con el método científico, para aceptar un enunciado es necesario plantear hipótesis, experimentar, y comprobarla con los resultados de los experimentos.

Cuando se trata de un experimento aleatorio no es fácil corroborar la hipótesis, debido a que se presentan diferentes resultados en cada ensayo y, por este motivo, se hace necesario un procedimiento especial para los fenómenos aleatorios. Por ejemplo, si un técnico de la compañía

de luz asegura que, en promedio, el voltaje que reciben los usuarios en cierta zona de la ciudad es de 117 voltios, resulta prácticamente imposible verificar con certeza esta afirmación.

La prueba de hipótesis es otra forma de inferencia que se utiliza para rechazar o aceptar hipótesis sobre un fenómeno aleatorio, planteada en términos de un parámetro o combinación de dos o más parámetros; también se usa para investigar si una distribución empírica se ajusta a un determinado modelo probabilístico.

Una hipótesis estadística es una afirmación o conjetura acerca de una o más poblaciones.

Por la naturaleza aleatoria de los fenómenos, la prueba de hipótesis estadística no se puede considerar como una prueba definitiva, ya que en ella se establece un criterio o regla de decisión para aceptar o rechazar la hipótesis y, como se verá posteriormente, la regla de decisión está asociada con ciertos riesgos.

### Conceptos básicos para la prueba de hipótesis

Una vez que una muestra ha sido tomada, generalmente esta se caracteriza por uno o más estadísticos muestrales. El propósito de la prueba de hipótesis es utilizar estos estadísticos y nuestro conocimiento de las distribuciones estadísticas para realizar inferencias acerca de la población de la cual la muestra se ha originado.

Las inferencias se pueden ser hechas acerca de la función de densidad de probabilidad de la población o del valor de parámetros poblacionales.

Una hipótesis es una sentencia estadística que puede ser aceptada o rechazada. Las hipótesis pueden ser formuladas para medias, variancias, diferencias de medias, o funciones de distribuciones de densidad de probabilidad. Algunos ejemplos son:

Prueba estadística	Descripción de hipótesis	Sentencia
Media	La media de la población es 25	$\mu = 25$
Dos medias	Dos poblaciones tienen la misma media	$\mu_1 = \mu_2$
Proporción	Fracción de población que es defectuosa	$\pi = 0.5$
FDP	Población sigue FDP de Poisson con $\lambda = 0.5$	$P(x; 0.5)$ es
Dos variancias	Dos poblaciones tienen variancias iguales	$\sigma_1^2 = \sigma_2^2$

Figura 3.2 Ejemplos de hipótesis

Existen dos hipótesis para cualquier prueba estadística. La primera y más importante es  $H_0$ , la hipótesis nula. La hipótesis alterna  $H_1$  es automáticamente aceptada si la prueba muestra que  $H_0$  debe de ser rechazada. Una hipótesis simple si esta involucra un solo valor, o compuesta si no proporciona un valor en específico. Una prueba de dos hipótesis simples es  $H_0: \mu = 25$  y  $H_1: \mu \neq 27$ . Una prueba de una simple y una compuesta es  $H_0: \mu = 25$  y  $H_1: \mu > 25$ .

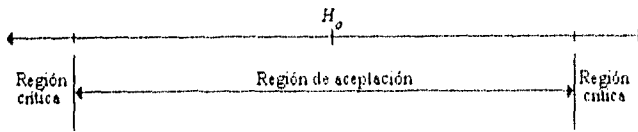


Figura 3.3 Regiones críticas y de aceptación para una prueba estadística

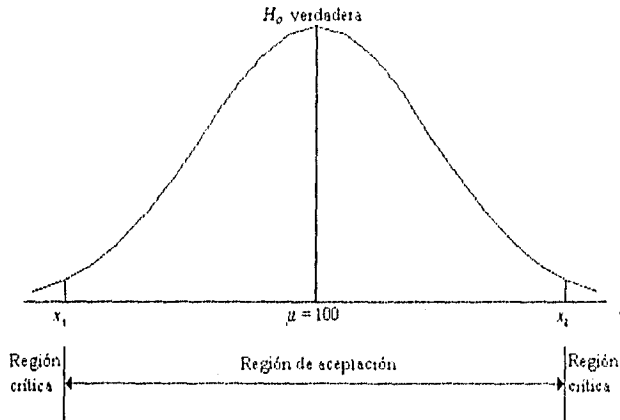


Figura 3.4 Regiones críticas y de aceptación para prueba de medias de una población normal

El deseo generalizado es no rechazar  $H_0$ . El término "no rechazar" es adecuado porque siempre es posible reformular la prueba de manera que se rechace  $H_0$ . Sin embargo para una fácil interpretación también se utilizara el término aceptar en vez de no rechazar  $H_0$ . Para aceptar  $H_0$ , el resultado de la prueba estadística debe de ser un número que caiga en la región de aceptación. Cualquier otro valor estará en la región crítica, como se muestra en la Figura 3.3, e implicará el rechazo de  $H_0$ . Por ejemplo, si la media de una población normal es  $\mu=100$  y se genera la hipótesis  $H_0: \mu = 100$  ó  $H_1: \mu \neq 100$ , dos valores  $x_1$  y  $x_2$  deben de ser determinados para definir las regiones críticas y de aceptación.

### Tipos de errores y sus probabilidades

Para  $H_0$  o  $H_1$  se pueden identificar dos tipos de error en nuestra decisión. Estos errores,

llamados tipo I y tipo II, son explicados en la Figura 3.5:

	Decisión	
Realidad	Rechazar $H_0$	Aceptar $H_0$
$H_0$ verdadera	Error tipo I	Correcto
$H_0$ falso	Correcto	Error tipo II

Figura 3.5 Tipos de errores

La probabilidad de estos errores es importante, puesto que determina que tan sensible y poderosa es la prueba en la distinción entre hipótesis verdaderas y falsas.

$$P(\text{Error tipo I}) = \alpha$$

$$P(\text{Error tipo II}) = \beta$$

### Prueba unilateral y prueba bilateral

Los términos significancia y área de confianza utilizados en la estimación por intervalos también aplican a la prueba de hipótesis. La significancia está indicada por  $\alpha$ , que es la probabilidad del error de tipo I. La confianza en la conclusión de la prueba es  $1 - \alpha$ . Entonces

$$\text{Significancia} + \text{confianza} = \alpha + (1 - \alpha) = 1$$

Generalmente, el nivel de  $\alpha$  se asigna previamente antes de ejecutar la prueba. Ejemplos son 0.10, 0.05, y 0.01, siendo 0.05 el más común. Si  $\alpha = 0.05$ , existe un nivel del 5% de significancia, lo cual significa que en promedio un error del tipo I (rechazar una  $H_0$  verdadera) ocurrirá 5 de cada 100 veces que sean probadas  $H_0$  y  $H_1$ . Además, existe un nivel de confianza del 95% de que el resultado sea correcto.

Si  $H_1$  involucra una relación no igual, sea  $H_1: \mu = 3.5$  y  $H_1: \mu \neq 3.5$ , sin ninguna indicación, entonces el área de significancia es dividida en partes iguales, una a cada lado de la distribución. Esta es llamada prueba bilateral o prueba de dos colas. Si es conocido que el parámetro se orientará a una sola dirección,  $H_0: \mu > 3.5$  o  $H_1: \mu < 3.5$  pueden ser utilizadas. Esta es llamada una prueba unilateral, y el área  $\alpha$  está en un solo lado de la distribución, con un solo lado. Puesto que la región crítica variará para prueba unilateral que para bilateral, es importante que las  $H_0$  y  $H_1$  apropiadas sean determinadas antes de que se realice la prueba.

### Procedimiento de la prueba de hipótesis

Por la naturaleza aleatoria de los fenómenos, la prueba de hipótesis estadística no se puede considerar como una prueba definitiva, ya que en ella se establece un criterio o regla de decisión

para aceptar o rechazar una hipótesis y, como se verá posteriormente, la regla de decisión está asociada con ciertos riesgos.

En primer lugar, para realizar una prueba de hipótesis estadística se debe de plantear el enunciado en cuestión, en términos de algún o algunos parámetros poblacionales como la media, la variancia, etc.

Generalmente se habla de dos tipos de hipótesis relacionadas con una prueba: la hipótesis nula ( $H_0$ ) y las hipótesis alternas ( $H_1, H_2, \dots$ ).

La hipótesis nula es la que se toma como referencia en la prueba y, por conveniencia, se plantea como una igualdad de la forma  $\theta = \theta_0$ , donde  $\theta_0$  es un valor supuesto del parámetro  $\theta$ .

Las hipótesis alternas son aquellas que difieren de  $H_0$  y pueden tener las siguientes formas:

$$\theta \neq \theta_0 \quad \theta > \theta_0 \quad \theta < \theta_0$$

Es importante señalar, que aceptar como verdadera una hipótesis alterna equivale a rechazar la hipótesis nula correspondiente y cuando se acepta  $H_0$ , se rechaza como verdadera  $H_1$ , considerando que es la única alternativa. Lo más importante de una prueba estadística es el diseño de una regla de decisión, la cual señala en que casos se debe aceptar o rechazar la hipótesis nula en función de que lo indica un estadístico, obtenido de una muestra representativa.

Para realizar inferencia estadística, el siguiente procedimiento estándar se sigue después de haber colectado los datos. Los pasos son:

1. Establecer la hipótesis. Desarrollar la forma exacta de  $H_0$  y  $H_1$ . La hipótesis alterna  $H_1$  es usada para hacer la prueba unilateral o bilateral.

2. Seleccionar el nivel de significancia. El valor de  $\alpha$  determinará la probabilidad de el error tipo I en la prueba.

3. Calcular los estadísticos muestrales y estimar los parámetros. Uno o más estadísticos pueden ser requeridos para ejecutar la prueba.

4. Calcular el estadístico de prueba. La fórmula utilizada para probar  $H_0$  es llamado estadístico de prueba. Esto resultará en un valor específico en la distribución usada para ejecutar la prueba, como lo puede ser la Normal,  $t$ , etc. Este cálculo requerirá de que se asuman ciertos valores acerca de la población.

5. Determinar las regiones críticas y de aceptación del estadístico de prueba. Se utiliza el nivel  $\alpha$  y los parámetros estimados, como sea necesario para encontrar el valor para una prueba con uno o dos lados en la tabla de la distribución apropiada.

6. Rechazar o aceptar  $H_0$ . Si el valor de estadístico calculada (Paso 4) está dentro de la región de aceptación, se acepta  $H_0$ ; si está dentro de la región crítica, se rechaza  $H_0$ .

### Dependencia entre $\alpha$ , $\beta$ , y el tamaño de la muestra

Existe una relación distinta entre los dos valores de probabilidad  $\alpha$  y  $\beta$  y el tamaño de la muestra  $n$  para cualquier hipótesis. El valor de cualquiera de estos es encontrado utilizando una prueba estadística y dando valores a los otros dos. generalmente los valores de  $\alpha$  y  $n$  son los más importantes, por lo que son predeterminados y el valor de  $\beta$  no es controlado. A continuación se establecen los procedimientos para calcular  $\alpha$ ,  $\beta$ , o  $n$ .

## INFERENCIA ESTADÍSTICA PARA MEDIAS

### Tipos de pruebas para medias

Uno de los usos más comunes para la prueba de hipótesis es determinar si las medias de ciertas muestras son iguales, mayores que, o menores que otras o que poseen un valor específico. Todas estas pruebas son llevadas a cabo utilizando una forma del estadístico  $z$  o del estadístico  $t$ .

$$z = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \quad t = \frac{(\bar{X} - \mu)\sqrt{n}}{s}$$

En la figura 3.6 se proporciona el tipo de estadístico. El estadístico  $t$  es utilizado solo cuando  $\sigma$  es desconocido y la muestra es pequeña.

Descripción de la(s) muestra(s)	$\sigma$ poblacional conocida o desconocida	$n$ grande o pequeña	Tipo de estadístico
Uno	Conocido	Grande o pequeña	$z$
	Desconocido	Grande	$z$
	Desconocido	Pequeña	$t$
Dos, independientes	Conocido	Grande o pequeña	$z$
	Desconocido	Grande	$z$
	Desconocido	Pequeña	$t$
Dos, no independientes (pareadas)	Desconocido	Grande	$z$
	Desconocido	Pequeña	$t$

Figura 3.6 Clasificación de los datos para la prueba de medias

Para poder utilizar la prueba correcta, se deben de contestar las siguientes preguntas.

I. ¿Existe una o dos muestras de datos? Si existen dos muestras, ¿Los elementos provienen de una misma fuente? Si esto es cierto, las muestras no son independientes, son pareadas.

2. ¿Es conocida la desviación estándar o debe de ser estimada?
3. ¿El tamaño de la muestra es grande? Considerando que una muestra es grande cuando  $n \geq 30$ .

### Prueba de medias para una muestra con desviación estándar conocida

Cuando un valor específico de media poblacional  $\mu_0$  es conocido o deseado, una media  $\mu$  estimada puede ser probada utilizando la hipótesis nula  $H_0: \mu = \mu_0$ . La hipótesis alterna  $H_1$  podría ser  $\mu \neq \mu_0$  ( bilateral ),  $\mu > \mu_0$  o  $\mu < \mu_0$  (unilateral). Ya sea que la desviación estándar sea conocida de datos históricos, o esta puede ser estimada por medio de  $\hat{\sigma}$  de una muestra grande ( $n \geq 30$ ) utilizada en esta prueba, en cuyo caso  $\hat{\sigma} = \sigma$ . El estadístico de prueba es la variable distribución Normal estándar para medias muestrales.

$$z = \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}$$

donde  $\mu$  es estimada del promedio  $\bar{X}$  de la muestra y que  $\mu_0$  es el valor establecido en  $H_0$ . Se utilizan, ahora, los pasos de la prueba de hipótesis.

La probabilidad  $\beta$  de un error de tipo II es dado por a curva de operación característica. La escala de las abscisas es diferente dependiendo de la forma de  $H_1$ .

$$\begin{aligned} \text{Bilateral, } H_1: \mu \neq \mu_0 & \quad d = \frac{|\mu - \mu_0|}{\sigma} \\ \text{Unilateral, } H_1: \mu > \mu_0 & \quad d = \frac{\mu - \mu_0}{\sigma} \\ \text{Unilateral, } H_1: \mu < \mu_0 & \quad d = \frac{\mu_0 - \mu}{\sigma} \end{aligned}$$

La curva de operación característica proporcionará un valor  $\beta$  para un tamaño  $n$  de la muestra. Si  $\beta$  es supuesta muy grande, se debe de tomar una muestra mayor y la prueba debe de volverse a hacer. Si  $\beta$  es seleccionada previamente a la colección de los datos, la curva de operación característica indica el valor requerido de  $n$ .

### Prueba de medias para una muestra con desviación estándar desconocida y muestra pequeña

Si  $\sigma$  es desconocida, como generalmente pasa, y el tamaño de la muestra es pequeño ( $n < 30$ ), el insesgado  $s = \hat{\sigma}$  es el mejor estimado de  $\sigma$ . Pero la distribución  $t$  es ahora utilizada para probar la hipótesis. La prueba estadística para una muestra es

$$t = \frac{(\mu - \mu_0)\sqrt{n}}{s}$$

donde  $\mu$  es estimado por el promedio de la muestra  $\bar{X}$  y  $\mu_0$  es el valor predeterminado para la media en  $H_0$ . La distribución  $t$  con  $v = n - 1$  grados de libertad es utilizada para establecer la región de aceptación.

Las curvas de operación característica para pruebas unilateral y bilateral con  $\alpha = 0.05$  son utilizadas para determinar  $\beta$ , o bien, el  $n$  requerido. Si se necesita un tamaño de muestra grande para asegurar un valor determinado de  $\beta$ , se puede utilizar la prueba distribución Normal estándar de la prueba de medias para una muestra con desviación estándar conocida.

### Prueba de medias para dos muestras independientes con desviación estándar conocida

Si las medias de dos muestras aleatorias independientemente inducidas desean ser probadas para igualdad o desigualdad y, las desviaciones estándar de ambas poblaciones son conocidas, la distribución Normal estándar es utilizada como estadístico de prueba. La prueba bilateral para muestras 1 y 2 es

$$H_0: \mu_1 = \mu_2 \quad \text{versus} \quad H_1: \mu_1 \neq \mu_2$$

La forma completa del estadístico de prueba es

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_d} \quad \text{Ecuación 3.1}$$

donde  $\sigma_d$  es la desviación estándar de la diferencia entre dos variables. La variancia de la diferencia es igual a la suma de las variancias individuales e independientes. Puesto que se está tratando con medias muestrales en la Ecuación 3.1,  $\sigma_d$  es la raíz cuadrada de la suma de las respectivas variancias de la media  $\sigma_i^2/n_i$  para  $i = 1, 2$ .

$$\sigma_d = \left( \frac{\text{varianza muestra 1}}{\text{tamaño muestra 1}} + \frac{\text{varianza muestra 2}}{\text{tamaño muestra 2}} \right)^{1/2} = \left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^{1/2} \quad \text{Ecuación 3.2}$$

Bajo  $H_0$  se tiene  $\mu_1 - \mu_2 = 0$ , por lo que la Ecuación 3.1 se reduce a la siguiente prueba estadística

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_d} \quad \text{Ecuación 3.3}$$

donde los valores  $\bar{X}$  son utilizados como los mejores estimadores de la media poblacional.

Si los valores de  $\sigma_i$  son desconocidos pero  $n_1$  y  $n_2$  exceden 30, las muestras son lo suficientemente grandes para estimar confiablemente  $\sigma_d$  utilizando los valores de la muestra  $s_i$ . La Ecuación 3.3 anterior sigue siendo utilizada como estadístico de prueba con  $s_i = \hat{\sigma}_i = \sigma_i$  en la Ecuación 3.2.



### Prueba de medias para dos muestras independientes con desviación estándar desconocida y muestras pequeñas

Si las desviaciones estándar poblacionales  $\sigma_1$  y  $\sigma_2$  son desconocidas, es necesario asumir que estas son, ya sea, iguales o diferentes y estimarlas de las muestras. Es posible determinar si estas son iguales por medio de la prueba  $F$  antes de ejecutar esta prueba para medias.

Si se asume igualdad y las muestras son pequeñas ( $n_1$  y  $n_2 < 30$ ), la distribución  $T$  es utilizada para probar  $H_0: \mu_1 = \mu_2$  contra la hipótesis alterna: El estadístico de prueba  $t$  y los grados de libertad son:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_d} \quad v = n_1 + n_2 - 2$$

El valor de  $s_d$  es la desviación estándar de la diferencia de medias. La variancia  $s_d^2$  se calcula como

$$s_d^2 = \frac{\text{suma combinada de los cuadrados de los residuales } n_1 + n_2}{\text{grados totales de libertad } n_1 n_2}$$

$$= \frac{\sum (X_{j1} - \bar{X}_1)^2 + \sum (X_{j2} - \bar{X}_2)^2}{(n_1 - 1) + (n_2 - 1)} \cdot \frac{n_1 + n_2}{n_1 n_2}$$

Ecuación 3.4

donde  $X_{j1}$  ( $j = 1, 2$ ) son valores observados para la muestra  $j$ . Es correcto substituir las variancias inesgadas de la muestra en el numerador de la Ecuación 3.4 puesto que

$$\sum (X_{j1} - \bar{X}_1)^2 + \sum (X_{j2} - \bar{X}_2)^2 = s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)$$

Si las variancias poblacionales son diferentes la prueba debe utilizar una distribución  $T$  aproximada. Si no existe una razón válida para asumir que  $\sigma_1^2 = \sigma_2^2$ , esta prueba debe ser utilizada. El estadístico de prueba y los grados de libertad son

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\left( \frac{s_1^2/n_1 + s_2^2/n_2}{2} \right)^{1/2}} \quad v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left( \frac{s_1^2/n_1}{n_1 + 1} + \frac{s_2^2/n_2}{n_2 + 1} \right)^2} - 2$$

Las curvas de operación característica de la prueba  $T$  no son confiables para esta prueba.

### Prueba de medias para muestras pareadas

Si la igualdad entre las medias de dos muestras es probada, cualquier diferencia que se presente puede ser atribuida a otros factores en particular. Por ejemplo, supóngase que productos equivalentes son fabricados en dos diferentes plantas y las medias de conductividad eléctrica son obtenidas para verificar su igualdad. Si las medias son estadísticamente diferentes, una razón puede ser la diferencia en el proceso. Sin embargo, podría ser por otros aspectos del proceso -la

maquinaria, nivel de capacitación, o las materias primas- pueden afectar también la diferencia. Para probar la diferencia de medias resultante de ciertos tratamientos, como el de los dos procesos, se hace con elementos provenientes de la misma fuente a los cuales se les da el tratamiento y después los resultados son medidos. De aquí que los datos serán organizados en muestras pareadas. El planteamiento es llevado a cabo mediante la división de un elemento en dos piezas o, en el caso de personas, utilizando gemelos como sujetos de prueba.

La prueba descrita aquí es solo para datos pareados, esta no debe ser utilizada para probar medias de dos muestras independientes.

La prueba  $t$  para datos pareados es usada para probar la hipótesis nula  $H_0: \delta = \delta_0$ , donde  $\delta$  es la diferencia de medias entre las muestras. El valor  $\delta_0$  es el valor hipotético de  $\delta$ , y generalmente  $\delta_0 = 0$ ; esto implica, que no hay diferencia entre las muestras que haya sido causada por los tratamientos. La media de las diferencias es  $\bar{d}$ , la cual es usada para estimar  $\delta$ . Para obtener  $\bar{d}$  se calculan las diferencias  $d_i = (i = 1, 2, \dots, n)$  de los datos  $X_{i1}$ , y  $X_{i2}$ .

$$d_i = X_{i1} - X_{i2}$$

$$\bar{d} = \frac{\sum d_i}{n}$$

El estadístico de prueba  $t$  es

$$t = \frac{(\bar{d} - \delta_0)\sqrt{n}}{s_d}$$

Ecuación 3.5

donde  $\bar{d} = \delta$ . La desviación estándar de la diferencia  $s_d$  es desconocida, por lo que se calcula con la fórmula insesgada para la desviación estándar

$$s_d = \left( \frac{\sum d_i^2}{n-1} - \frac{n}{n-1} \bar{d}^2 \right)^{1/2}$$

El procedimiento de prueba de hipótesis es utilizado para esta prueba.

Los grados de libertad para la distribución  $t$  son

$$\nu = n - 1$$

puesto que solo hay  $n$  elementos diferentes de la muestra. Esto significa que la precisión de la prueba  $t$  pareada es menor que la prueba de dos muestras donde  $\nu = n_1 + n_2 - 2$ . Por lo tanto la curva de operación característica mostrará un valor mayor de  $\beta$  para datos pareados.

Si el tamaño de la muestra es grande, el estadístico de prueba de la Ecuación 3.5, es utilizado, excepto si está en términos de  $z$  y la distribución Normal estándar se utiliza para definir la región de aceptación.

## INFERENCIA ESTADÍSTICA PARA VARIANCIAS

### Tipos de pruebas para variancias

En todas las pruebas para medias de dos muestras fue necesario hacer algunas asunciones acerca de la igualdad o desigualdad de valores de  $\sigma^2$ . Las pruebas de esta sección son utilizadas para determinar la relación entre las variancias de las dos poblaciones.

Todas las pruebas para las variancias utilizan ya sea  $\chi^2$  o el estadístico  $F$ .

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad F = \frac{s_1^2}{s_2^2}$$

Se asume que las muestras son tomadas de una población  $N(\mu, \sigma^2)$  donde  $\mu$  no necesariamente es conocida. Las siguientes preguntas son usadas para determinar el estadístico de prueba apropiado.

1. ¿Existen una o dos muestras de datos? Si es una, corresponde a la prueba de variancia para una muestra. Si son dos, corresponde a la prueba de variancia para dos muestras.
2. ¿Los datos provienen de una población normal? Si esto es cierto, se procede; en caso contrario, se requiere una  $n \geq 30$  para estimar variancias confiablemente.

### Prueba de variancias para una muestra

Si la variancia de una muestra debe ser igual a un valor específico de  $\sigma_0^2$ , la hipótesis nula es  $H_0: \sigma^2 = \sigma_0^2$  y cualquier uni o bilateral  $H_1$  se puede probar. El valor de  $\sigma_0^2$ , puede ser un valor estándar, históricamente observado, o un valor deseado.

El estadístico de prueba es la variable  $\chi^2$  con  $v = n - 1$  grados de libertad:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

La variancia insesgada de la muestra  $s^2$  es calculada por la siguiente ecuación

$$s^2 = \frac{\sum f_i X_i^2}{n-1} - \frac{n}{n-1} \bar{X}^2$$

El procedimiento de prueba de hipótesis es utilizado para ejecutar esta prueba.

### Prueba de variancias para dos muestras

Para probar la igualdad de variancias de dos poblaciones normales, el estadístico  $F$  es usado

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

Bajo la hipótesis nula  $H_0: \sigma_1^2 = \sigma_2^2$  el estadístico de prueba se reduce a

$$F = \frac{s_1^2}{s_2^2}$$

lo cual tiene la distribución  $F = (v_1, v_2)$ , donde

$$v_1 = n_1 - 1 \quad \text{y} \quad v_2 = n_2 - 1$$

Si se planea probar por la igualdad de dos medias muestrales con variancias iguales, el bilateral  $H_1: \sigma_1^2 \neq \sigma_2^2$  debe ser rechazada. La región de aceptación tiene  $\alpha/2$  del área en cada lado de la distribución  $F = (v_1, v_2)$ . Para determinar el límite izquierdo de la región de aceptación se usa la ecuación

$$F_0' = \frac{1}{F(v_2, v_1)}$$

Para una prueba unilateral, se define  $s_1^2$  como la variancia de la muestra más grande, y la hipótesis alterna es siempre  $H_1: \sigma_1^2 > \sigma_2^2$ . El estadístico de prueba es  $F \geq 1$ , por lo que

$$F = \frac{\text{variancia más grande}}{\text{variancia más pequeña}} = \frac{s_1^2}{s_2^2}$$

## INFERENCIA ESTADÍSTICA PARA PROPORCIONES

### Tipos de pruebas para proporciones

La aproximación Normal a la Binomial es generalmente precisa si  $np > 5$  y  $nq < 5$ . Esta guía general también aplica a la prueba de proporciones de poblaciones para una o dos muestras utilizando el estadístico normal

$$z = \frac{\text{valor} - \text{media}}{\text{desviación estándar}}$$

Si las muestras son demasiado pequeñas para usar la aproximación Normal, la distribución Binomial debe de ser usada.

Si la condición de la aproximación Normal no es violada, la única pregunta a contestar antes de probar una proporción es, ¿Existen una o dos muestras de datos? Si el caso que se presenta es el primero, se realiza la prueba de proporciones para una muestra, en el otro caso se realiza la prueba de proporciones para dos muestras.

### Prueba de proporciones para una muestra

Si una proporción muestral se supone que sea estadísticamente igual a un valor específico  $p_0$ , la hipótesis nula es  $H_0: \pi = p_0$ . Cualquiera de las hipótesis alternas unilaterales o bilaterales se puede probar.

El estadístico de prueba de aproximación Normal es

$$z = \frac{p - p_0}{\sigma_p}$$

Ecuación 3.6

donde  $p$  es calculado de la muestra y de

$$\sigma_p = \sqrt{\frac{p_0(1-p_0)}{n}}$$

puesto que bajo la hipótesis nula el valor verdadero de la proporción es  $p_0$ . Entonces, cada vez que un nuevo valor de  $\pi$  es hipotetizado para una población, la desviación estándar cambia.

Es fácil de desarrollar el estadístico de la Ecuación 3.6 para que sea probado el número en vez de la proporción. Puesto que la media y la desviación estándar para la distribución Binomial son  $np$  y  $\sqrt{np(1-p)}$ , respectivamente, el estadístico de prueba es

$$z = \frac{np - np_0}{\sqrt{np_0(1-p_0)}}$$

### Prueba de proporciones para dos muestras

Para determinar si dos valores de una proporción pertenecen a una misma población, la hipótesis nula es  $H_0: \pi_1 = \pi_2$ . El estadístico completo de prueba es

$$z = \frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sigma_{p_1 - p_2}}$$

donde  $\sigma_{p_1 - p_2}$  es la desviación estándar de las diferencias entre las dos proporciones. esta es una aproximación Normal a la Binomial y es precisa solo si  $np_i > 5$  y  $nq_i > 5$  para  $i = 1, 2$ . Bajo la hipótesis nula  $\pi_1 - \pi_2 = 0$ , el estadístico de prueba se reduce a

$$z = \frac{p_1 - p_2}{\sigma_{p_1 - p_2}}$$

El procedimiento para determinar el valor de  $\sigma_{p_1 - p_2}$  es diferente si  $\pi_1$  y  $\pi_2$  son conocidos en vez de ser estimados de la muestra de datos. Si los valores son conocidos con base en la experiencia, o por otro medio, se aplica la regla que dice que la variancia de la diferencia es la suma de las variancias si y solo si las variables son independientes, para calcular la variancia de la diferencia de dos variables independientes. Las variancias son sumadas para obtener

$$\sigma_{p_1 - p_2} = \left[ \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2} \right]^{1/2}$$

Si los valores de  $\pi_1$  y  $\pi_2$  son desconocidos, un estimado de  $\pi$  se obtiene mediante la combinación de dos muestras y el cálculo de  $p$ . entonces el estimado de  $\pi_1$  y  $\pi_2$  es  $s_{p_1 - p_2}$  donde  $q = 1 - p$ .

$$s_{p_1 - p_2} = \left( \frac{pq}{n_1} + \frac{pq}{n_2} \right)^{1/2} \quad s_{p_1 - p_2} = \left( pq \frac{n_1 + n_2}{n_1 n_2} \right)^{1/2}$$

Esta combinación de las muestras 1 y 2 es posible puesto que  $H_0$  hipotetiza que  $\pi_1 = \pi_2$ .

## INFERENCIA ESTADÍSTICA PARA BONDAD DE AJUSTE

### Tipos de pruebas para bondad de ajuste

En todas las inferencias estadísticas estudiadas se han involucrado parámetros poblacionales como media, variancia, y proporciones. Ahora se realizarán inferencias acerca de una completa distribución poblacional. Una muestra es tomada y se desea probar una hipótesis nula de la forma general

$H_0$ : la muestra es de una distribución especificada

Cualquier distribución puede ser probada, sin importar si esta es bien conocida (Binomial, Poisson, Normal, etc.) o de que sea desarrollada por el estadístico. La hipótesis alterna es siempre de la forma

$H_1$ : la muestra no es de la distribución especificada

Una prueba de  $H_0$  versus  $H_1$  es llamada prueba de bondad de ajuste.

Dos pruebas son utilizadas para evaluar la bondad de ajuste:

1. La prueba  $\chi^2$ , la cual está basada en la aproximación al estadístico  $\chi^2$ .

2. La prueba Kolmogorov-Smirnov (K-S). Esta es llamada una prueba no paramétrica puesto que esta usa un estadístico de prueba que no hace asunciones acerca de la distribución.

El estadístico de prueba  $\chi^2$  es la mejor opción para probar distribuciones discretas, y la prueba K-S es la mejor para distribuciones continuas para las cuales los valores de los parámetros son especificados independientemente de la muestra de datos. La prueba  $\chi^2$  es usada para evaluar cuantitativamente el ajuste para datos observados, lo cual se realiza de forma subjetiva en la Binomial y la de Poisson. Para determinar que prueba usar, se deben de contestar las siguientes preguntas.

1. ¿La distribución a ser probada es discreta o continua?
2. ¿Son especificados los valores de los parámetros para la distribución?

Si no son especificados, estos deben ser estimados de la muestra.

Distribución	Parámetros especificados	Prueba a utilizar
discreta o continua	o estimados	utilizar
Discreta	Especificado	$\chi^2$
Discreta	Estimado	$\chi^2$
Continua	Especificado	K-S
Continua	Estimado	$\chi^2$

Figura 3.4 Pruebas apropiadas para bondad de ajuste

Es importante entender que la prueba de bondad de ajuste no puede ser utilizada para seleccionar la mejor distribución. Después de que se ha elegido la distribución, la prueba utiliza los datos observados para determinar estadísticamente si la opción elegida fue la correcta.

### $\chi^2$ Prueba de bondad de ajuste para distribuciones discretas

La prueba  $\chi^2$  utiliza un estadístico que tiene una distribución aproximada a la  $\chi^2$ . Esta es una medida de la diferencia relativa entre los valores de frecuencia observados y esperados para cada valor de variable

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad \text{Ecuación 3.7}$$

donde  $k$  = número de valores diferentes de variable

$O_i$  = valor de frecuencia observada

$E_i$  = valor esperado o teórico de la frecuencia

Los valores  $E_i$  son el producto del tamaño de la muestra y la probabilidad de la distribución hipotetizada.

$$E_i = nP_i \quad \text{Ecuación 3.8}$$

Puesto que la función de densidad de probabilidad  $\chi^2$  es usada para probar el ajuste, se deben de determinar los grados de libertad. Esto es

$$v = k - r - 1$$

donde  $r$  es el número de parámetros de la distribución hipotetizados que es estimada de la muestra de datos.

Los pasos de la prueba de hipótesis son usados para la prueba de bondad de ajuste. Los valores  $E_i$  y el de  $\chi^2$  es calculado en el paso 4, de dicha prueba. Esta prueba es siempre una prueba unilateral, entonces la región de aceptación para  $H_0$  son todos los valores de  $\chi^2$  debajo de algunos valores de  $\chi^2_0$  con un área a la izquierda de  $1 - \alpha$ ; esto es

$$P(\chi^2 \leq \chi^2_0) = 1 - \alpha$$

### $\chi^2$ Prueba de bondad de ajuste para distribuciones continuas

El procedimiento para probar cualquier distribución continua es idéntico al anteriormente presentado, excepto que los datos deben ser agrupados en  $k$  celdas antes de que los valores de  $E_i$  sean calculados para una distribución hipotética. Deben de existir por lo menos cinco celdas, y la probabilidad  $P_i$  de que la variable esté en la celda  $i$  es usada para calcular con la Ecuación 3.8. Esto es a menudo usado para determinar si los datos son de una distribución Normal una vez que los valores esperados son calculados por los pasos de ajuste de la Normal a los datos observados.

□

**CICLO DE VIDA DE LOS SISTEMAS**

□ El paradigma del ciclo de vida demanda un enfoque sistemático y secuencial para desarrollar software, que progresa por el análisis, el diseño, la codificación, la prueba y el mantenimiento.

Modelado, después de el ciclo de vida convencional para las disciplinas ingenieriles, el paradigma de el ciclo de vida comprende las siguientes actividades:

- Ingeniería de software y diseño: Puesto que el sistema de computadora es siempre parte de un sistema mayor, el trabajo comienza mediante la definición de requerimientos para todos los elementos del sistema y así proceder con la asignación de algunos de estos para el software. Este punto de vista del sistema es esencial cuando el sistema debe comunicarse con otros elementos.
- Análisis de requerimientos del sistema: La recopilación de requerimientos se intensifica y se enfoca específicamente al software. Para entender la naturaleza de los programas a ser construidos, el ingeniero de software debe entender el dominio de la información en la que el sistema se desenvolverá, así como las funciones requeridas, el desempeño, y las interfaces. Como medida de suma importancia, los requerimientos para el sistema deben ser revisados junto con el cliente.
- Diseño: El diseño del sistema es un proceso de pasos múltiples que se enfoca en tres distintos atributos de los programas: estructura de datos, arquitectura del software, y detalle procedural. El proceso de diseño traduce los requerimientos en una representación de software que puede ser evaluada antes de ser codificada. Así como los requerimientos, el diseño se documenta y se convierte en parte de la configuración del software.



- **Codificación:** El diseño debe de ser traducido en una forma entendible por la máquina. El paso de codificación ejecuta esta tarea. Si el diseño es realizado de una manera detallada, la codificación puede llevarse a cabo de una manera casi mecánica.
- **Prueba:** Una vez que la codificación se ha generado, la prueba del programa comienza. El proceso de prueba se enfoca en el interior lógico del software, asegurando que todas los planteamientos han sido probados y, en los exteriores funcionales, esto es, conducir la prueba para asegurar que una entrada definida producirá una salida que coincidirá con lo esperado.
- **Mantenimiento:** El software, indudablemente, sufrirá cambios después de que ha sido liberado al cliente. Los cambios ocurrirán puesto que se detectarán errores a consecuencia de la adaptación del software en el ambiente externo, o a causa de que el cliente requiere de un cambio funcional, o bien de adecuación del rendimiento.

El paradigma del ciclo de vida tiene un definitivo e importante lugar en la ingeniería de software. Este provee un esquema en el cual los métodos para el análisis, el diseño, la codificación, la prueba y el mantenimiento pueden llevarse a cabo.

El ciclo de vida de los sistemas permanece como el modelo procedural para la ingeniería de software más ampliamente utilizado.

## **ANÁLISIS**

El análisis de sistemas es una actividad comprendida dentro de la ingeniería de software. El análisis de sistemas se enfoca a todos los elementos del sistema, no solo a los del software.

El análisis de sistemas es conducido por los siguientes objetivos:

1. Identificar las necesidades del cliente.
2. Evaluar la factibilidad conceptual del sistema.

### **Identificación de las necesidades**

El primer paso en el proceso de análisis de sistemas involucra la identificación de las necesidades.

La identificación de las necesidades es el punto de inicio en la evolución de un sistema basado en computadora. Para comenzar, el ingeniero se reúne con el cliente y lo asiste en la definición de metas del sistema, la información a ser provista, la información a ser producida, las funciones y el rendimiento requerido. El ingeniero debe asegurarse de distinguir entre las necesidades del cliente (aspectos críticos para el éxito) y los deseos del cliente (aspectos deseables a tener, pero no esenciales).

Después de un par de entrevistas con el usuario y de haber discutido en general el sistema, se definieron las siguientes necesidades:

- Desarrollar un sistema para computadora que en forma interactiva guíe al usuario para realizar las pruebas estadísticas más usuales.
- La interfaz del sistema con el usuario debe de ser amigable. El sistema debe de presentar opciones en ventanas tipo menú.
- Debe ejecutarse sobre el ambiente Excel de Microsoft, por lo que se recomienda utilizar macros para programarlo.

- Debe contener:

Parámetros básicos

Estadística descriptiva

Modelos probabilísticos

Binomial

Exponencial

Normal

Pascal

Poisson

Pruebas de bondad de ajuste

$\chi^2$  Binomial

$\chi^2$  Discreta

$\chi^2$  Normal

$\chi^2$  Poisson

Pruebas de hipótesis

Igualdad de medias

Una muestra desviación estándar conocida

Una muestra desviación estándar desconocida

Dos muestras desviación estándar conocida

Dos muestras desviación estándar desconocida

Muestras pareadas

Igualdad de variancia

Una muestra

Dos muestras

Igualdad de proporciones

Una muestra

Dos muestras

- El usuario debe de proporcionar datos de entrada y deberá obtener la salida ordenada y organizada, esta puede ser presentarse en un menú o impresa en una hoja de cálculo común.

Una vez que las necesidades han sido definidas, comienza la evaluación de toda la información obtenida.

### **Estudio de factibilidad**

Todos los proyectos son factibles, si es que se les proporciona recursos ilimitados y tiempo infinito. Desafortunadamente, los desarrollos de sistemas para computadora están acosados por la escasez de recursos y la presión de fechas de entrega comprometidas. En ambos casos es prudente evaluar la factibilidad del proyecto en el menor tiempo posible.

El estudio de factibilidad se concentra en cuatro áreas principales de interés:

*Factibilidad económica:* Evaluar el costo que se va a ejercer contra el beneficio que se espera obtener.

Hay muchas ocasiones en que el beneficio que se obtiene de una inversión no cubre el gasto en que se ha incurrido. Esto se justifica siempre y cuando los beneficios intangibles que provengan del proyecto no sean cuantificables o medibles con exactitud, pero que sean definitivamente estratégicos o vitales. Este es el caso de este trabajo, ya que lo que se busca es desarrollar un proyecto que sirva de base para el desarrollo del trabajo escrito de una tesis profesional.

*Factibilidad técnica:* Estudiar la posibilidad de alcanzar un sistema aceptable, en cuanto a funcionalidad, desempeño y posibles restricciones.

El desarrollo de este sistema propone una gran ventaja y otra desventaja. La ventaja es que se utilizan las facilidades tanto de desarrollo de interfaces gráficas como de lenguaje de programación de macros en Excel. La desventaja es que la programación de macros exige una codificación elaborada, semejante a la de lenguaje ensamblador.

Sin embargo el balance de ambas arroja un saldo positivo y conveniente, que garantiza la elaboración de un sistema acorde a las necesidades del cliente.

### **Análisis de Requerimientos**

El análisis de requerimientos es la tarea de la ingeniería de software que sirve de puente entre los niveles de asignación de software y el diseño de software. El análisis de requerimientos habilita para especificar la función y el desempeño del software, indicando las interfaces del software con otros elementos del sistema, y estableciendo las limitaciones de diseño que el software debe de acatar. El análisis de requerimientos permite definir el dominio de información donde va a desarrollarse el sistema. También, provee de una representación de la información en funciones que deben ser traducidas en datos, arquitecturas, y procedimientos. Finalmente, la especificación de los requerimientos provee a el desarrollador y a el cliente con medios para lograr construir software de calidad.

## Tareas del análisis

El análisis de requerimientos de sistemas puede ser dividido en cuatro áreas de esfuerzo:

1. Reconocimiento del problema.
2. Evaluación.
3. Especificación.

Inicialmente, el ingeniero de software estudia las especificaciones del sistema y, si existe, el plan del proyecto. Es importante entender el software en el contexto del sistema y revisar el alcance del sistema. Después, la comunicación para el análisis debe ser establecida para que el reconocimiento del problema sea asegurado.

La evaluación del problema y la síntesis de la solución es el área de mayor esfuerzo. Se debe de evaluar el flujo y la estructura de la información, depurar todas las funciones del software en detalle, establecer las características de las interfaces, e identificar las restricciones. Cada una de estas tareas sirve para describir el problema de tal manera que pueda ser sintetizado el enfoque de las soluciones.

El proceso de evaluación continua hasta que tanto el ingeniero de software y cliente se sienten confiados de que el sistema puede ser adecuadamente desarrollado.

Las tareas asociadas con el análisis existen para proveer una representación del sistema que pueda ser revisada y aprobada por el cliente. En un caso ideal el cliente desarrolla por completo la especificación de los requerimientos del software. Este es un caso verdaderamente inusual; la mejor especificación la realizan en conjunción ingeniero y cliente.

*Reconocimiento del problema:* El problema principal que atañe a este sistema, es que no existe una herramienta que facilite el uso o compile las funciones estadísticas más comúnmente usadas.

*Evaluación:* En conjunción con el usuario o cliente, se evaluaron las posibles pruebas estadísticas a incluir en el sistema. Se determinaron cuales y se determinó que es factible realizar un sistema que las conjunte.

*Especificación:* El sistema debe de ser desarrollado en computadora. El sistema debe de ser amigable y debe de incluir los aspectos detallados en la identificación de necesidades.

## DISEÑO

El diseño de software es un proceso a través del cual los requerimientos se traducen en alguna representación de software. Inicialmente la representación del sistema es muy general. Depuraciones subsecuentes originan el diseño de una representación muy cercana del código fuente.

Desde el punto de vista de manejo de proyectos, el diseño de software se conduce en dos pasos:

1. Diseño preliminar. El cual comprende la transformación de los requerimientos en datos y arquitectura de software.
2. Diseño detallado. Este se enfoca a depurar la representación arquitectónica, la cual contendrá estructuras de datos y representaciones algorítmicas detalladas.

### **Diseño preliminar**

Dada las características de las necesidades del cliente y las facilidades que brinda Microsoft Excel, el diseño preliminar es:

Se programaran siete módulos independientes, la razón de que sean independientes es que cada uno de estos deberá ser activado mediante el icono correspondiente, incluido en la barra de herramientas. Los nombre de estos módulos serán:

*FI.* Información referente a la tesis profesional.

*Básicos.* Parámetros básicos de estadística descriptiva.

*Modelos.* Modelos probabilísticos.

*Bondad.* Pruebas de bondad de ajuste.

*Medias.* Pruebas de hipótesis de medias.

*Varian.* Pruebas de hipótesis de variancias.

*Proporci.* Pruebas de hipótesis de proporciones.

La descripción de la arquitectura de estos módulos es la siguiente:

Todos los módulos deberán ser presentados una ventana de diálogo (*dialog box*) que pueda ser activada inicialmente al elegir un icono con el mouse. Cada una de estas ventanas de diálogo deberán presentar un menú con opciones que de manera amigable, formato Excel, permita navegar hasta que se proporcionen los datos requeridos para la prueba estadística a realizar. Los resultados, según el caso, deberán ser presentados como parte de la ventana de diálogo en que se proporcionaron los datos, o como parte de la hoja de cálculo donde se proporcionaron estos.

*FI:* Se desplegará la información correspondiente a la tesis, quien la presenta y el director.

*Básicos:* A a partir de una muestra, dada como datos de una hoja de cálculo, debe organizarla y ordenarla, además de la siguiente información y medidas de estadística descriptiva: Número de Elementos, Mínimo, Máximo, Rango, Media, Mediana, Moda, Variancia, Desviación Estándar, Curtosis y Asimetría. Debe, también, presentar la tabla de frecuencias que contenga Marcas de Clase, Frecuencias Relativas y Frecuencias Acumuladas.

*Modelos:* Dados los parámetros requeridos por cada uno de los modelos probabilísticos, este módulo, deberá proporcionar las funciones de probabilidad de cada uno de ellos. Los modelos

elegidos son: Binomial, Exponencial, Normal, Pascal y Poisson. La entrada-salida de la información deberá ser a través de ventanas diálogo.

*Bondad:* Dada las muestras en una hoja de cálculo, se realizarán las Pruebas de Bondad de Ajuste Chi Cuadrada siguientes: Binomial, Discreta, Normal y Poisson. La información de salida deberá estar organizada y contener la muestra, el número de elementos, las probabilidades observadas y esperadas, el valor de  $\chi^2$ , grados de libertad, nivel de significancia<sup>3</sup>, el valor de  $\chi_0^2$ , las hipótesis  $H_0$  y  $H_1$ , y el resultado de la hipótesis. En los casos que sea necesario se proporcionará información adicional. Los resultados deberán quedar como parte de la hoja de cálculo original.

*Medias:* Dada una muestra, o dos según el caso, e información adicional propia de la prueba, se deberán realizar las siguientes Pruebas de Hipótesis: Medias para Una Muestra con Desviación Estándar Conocida, Medias para Una Muestra con Desviación Estándar Desconocida y Muestra Pequeña, Medias para Dos Muestras Independientes con Desviación Estándar Conocida, Medias para Dos Muestras Independientes con Desviación Estándar Desconocida y Muestras Pequeñas y Medias para Muestras Pareadas. Se requiere la siguiente información: muestra(s), ocurrencias, unilateral/bilateral, media poblacional y muestral, grados de libertad, nivel de significancia, las hipótesis  $H_0$  y  $H_1$ , y el resultado de la hipótesis. En los casos que sea necesario se proporcionará información adicional. Los resultados deberán quedar como parte de la hoja de cálculo original.

*Variancias:* Dada una muestra, o dos según el caso, e información adicional propia de la prueba, se deberán realizar las siguientes Pruebas de Hipótesis: Variancias para Una Muestra y Variancias para Dos Muestras. Se requiere la siguiente información: muestra(s), ocurrencias, unilateral/bilateral, variancia poblacional y muestral, grados de libertad, nivel de significancia, las hipótesis  $H_0$  y  $H_1$ , y el resultado de la hipótesis. En los casos que sea necesario se proporcionará información adicional. Los resultados deberán quedar como parte de la hoja de cálculo original.

*Proporciones:* Dada una muestra, o dos según el caso, e información adicional propia de la prueba, se deberán realizar las siguientes Pruebas de Hipótesis: Proporciones para Una Muestra y Proporciones para Dos Muestras. Se requiere la siguiente información: muestra(s), ocurrencias, unilateral/bilateral, proporción histórica y muestral, grados de libertad, nivel de significancia, las hipótesis  $H_0$  y  $H_1$ , y el resultado de la hipótesis. En los casos que sea necesario se proporcionará información adicional. Los resultados deberán quedar como parte de la hoja de cálculo original.

### Diseño detallado

A continuación se incluyen los pseudo códigos en los que se basaron los siete módulos del

---

<sup>3</sup> El nivel de significancia será siempre 0.05.

sistema. Se deberán agregar rutinas de entrada, validación y de formateo de información. Adicionalmente, es de esperarse que la codificación se vaya adecuando a las facilidades del lenguaje de programación.

### **Información**

#### **Rutina\_FI**

Despliega(Facultad de Ingeniería UNAM)  
Despliega(Tesis Profesional Ingeniería en Computación)  
Despliega(Sistema de Apoyo a Toma de Decisiones en Inferencia Estadística)  
Despliega(Presentada por: José Luis Pichardo López)  
Despliega(Director: M.I. Victor Flores Zavala)

#### **Fin\_Rutina\_FI**

### **Parámetros Básicos de Estadística Descriptiva**

#### **Rutina\_Rango**

Entra\_Arreglo(*Inicial, Final, Arreglo*)  
Ordena\_Arreglo(*Arreglo, Mayor, Menor, Arreglo\_Ordenado*)  
Rango = Mayor-Menor  
Despliega(Rango)

#### **Fin\_Rutina\_Rango**

#### **Rutina\_Frecuencias**

Obten\_Frecuencias(*Arreglo\_Ordenado, Arreglo\_Frecuencias*)  
Despliega(*Arreglo\_Frecuencias*)

#### **Fin\_Rutina\_Frecuencias**

#### **Rutina\_Frecuencias\_Relativas**

Frecuencia\_Acumulada = 0  
**FOR** I = 1 **TO** Numero\_de\_Intervalos **DO**  
    Frecuencia\_Relativa[I] = Arreglo\_Frecuencias[I]/Numero\_de\_Intervalos  
    Frecuencia\_Acumulada = Frecuencia\_Acumulada + Frecuencia\_Relativa[I]  
    Despliega(Frecuencia\_Relativa[I])

#### **ENDFOR**

Despliega(Frecuencia\_Acumulada)

#### **Fin\_Rutina\_Frecuencias\_Relativas**

#### **Rutina\_Media**

Obten\_Media(*Arreglo\_Ordenado, Media*)  
Despliega(Media)

#### **Fin\_Media**

#### **Rutina\_Mediana**

Cuenta(*Arreglo\_Ordenado, Cuantos\_Elementos*)  
Divide(*Cuantos\_Son, 2, Cociente, Residuo*)

```

IF Residuo = 0 THEN
    Mediana = (Arreglo_Ordenado[Cociente]+Arreglo_Ordenado[Cociente+1])/2
ELSE
    Mediana = (Arreglo_Ordenado[Cociente])
ENDIF
Despliega(Mediana)
Fin_Rutina_Mediana

```

```

Rutina_Variancia
    Obten_Variancia(Arreglo_Ordenado,Variancia)
    Despliega(Variancia)
Fin_Rutina_Variancia

```

```

Rutina_Desviacion_Estandar
    Desviacion_Estandar =  $\sqrt{\text{Variancia}}$ 
    Despliega(Desviacion_Estandar)
Fin_Rutina_Desviacion_Estandar

```

```

Rutina_Coeficiente_de_Variacion
    Coeficiente_de_Variacion = Desviacion_Estandar/Media
    Despliega(Coeficiente_de_Variacion)
Fin_Rutina_Coeficiente_de_Variacion

```

#### Modelos Probabilísticos

```

Rutina_Binominal
    Obten_Binominal(Exitos_Binominal,Ensayos_Binominal,Masa_Binominal,Prob_Binominal)
    Despliega(Prob_Binominal)
Fin_Rutina_Binominal

```

```

Rutina_Exponencial
    Obten_Exponencial(Periodo_Expo,Promedio_Expo,Masa_Expo,Prob_Expo)
    Despliega(Prob_Expo)
Fin_Rutina_Exponencial

```

```

Rutina_Normal
    Obten_Normal(Equis_Normal,Media_Normal,DE_Normal,Masa_Normal,Prob_Normal)
    Despliega(Prob_Normal)
Fin_Rutina_Normal

```

```

Rutina_Pascal
    Obten_Pascal(Exitos_Pascal,Ensayos_Pascal,Prob_Pascal)
    Despliega(Prob_Pascal)
Fin_Rutina_Pascal

```



### **Rutina\_Poisson**

Obten\_Poisson(*Eventos\_Poisson, Promedio\_Poisson, Masa\_Poisson, Prob\_Poisson*)  
Despliega(*Prob\_Poisson*)

**Fin\_Rutina\_Poisson**

### **Pruebas de Bondad de Ajuste**

#### **Rutina\_Chi2\_Binomial**

Entra\_Arreglo(*Inicial, Final, Ensayos*)

Entra\_Arreglo(*Inicial, Final, Exitos\_X\_Ensayo*)

Entra(*Ene*)

Entra(*Probabilidad*)

k = Ene

Cuenta(*Ensayos, Cuantos\_Ensayos*)

Cuenta(*Exitos\_X\_Ensayo, Cuantos\_EXE*)

Sumatoria(*Exitos\_X\_Ensayo, Sumatoria\_EXE*)

**IF** Cuantos\_Ensayos = Cuantos\_EXE **THEN**

Binomial[0] = 0

**FOR** i = 1 **TO** Cuantos\_Ensayos **DO**

Binomial[i] = Obten\_Binomial(*Ensayos[i], Ene, Probabilidad*)

**IF** i ≤ Ene **THEN**

Binomial\_SS[i] = Binomial[i] - Binomial[i-1]

**ELSE**

Binomial\_SS[i] = 1 - Binomial[i]

**ENDIF**

Esperada[i] = Sumatoria\_EXE \* Binomial\_SS[i]

**ENDFOR**

Indice\_CFM = Cuantos\_Ensayos

**WHILE** Indice\_CFM ≠ 0 **DO**

**IF** Esperada[Indice\_CFM] < 5 **THEN**

Acumulada\_CFM = 0

Acumulada\_O\_CFM = 0

**WHILE** Acum\_CFM < 5 **OR** Indice\_CFM ≥ 0 **DO**

Acumulada\_CFM = Acumulada\_CFM + Esperada[i]

Acumulada\_O\_CFM = Acumulada\_O\_CFM + Exitos\_X\_Ensayo[i]

Esperada[Indice\_CFM] = Acumulada\_CFM

Exitos\_X\_Ensayo[Indice\_CFM] = Acumulada\_O\_CFM

Indice\_CFM = Indice\_CFM - 1

**ENDWHILE**

**ELSE**

Indice\_CFM = Indice\_CFM - 1

**ENDIF**

**ENDWHILE**

Chi2\_Acumulada = 0

**FOR** i = 1 **TO** Cuantos\_Periodos **DO**

```

IF Esperanza[1] > 5 THEN
    Chi2[1] = (Exitos_X_Ensayo[1] - Esperada[1])2 / Esperada[1]
    Chi2_Acumulada = Chi2_Acumulada + Chi2[1]
    k = I
ENDIF
ENDFOR
Grados_libertad = k - 1
Chi2_Cero = Obten_Chi2(Grados_Libertad, 0.05)
Despliega("El valor de  $\chi^2$  es" Chi2_Acumulada)
Despliega("El valor de  $\chi_0^2$  es" Chi2_Cero)
IF Chi2_Acumulada ≤ Chi2_Cero THEN
    Despliega(" $\chi^2 \leq \chi_0^2$ ")
    Despliega("Por lo tanto se acepta  $H_0$ ")
ELSE
    Despliega(" $\chi^2 > \chi_0^2$ ")
    Despliega("Por lo tanto se rechaza  $H_0$ ")
ENDIF
ELSE
    Despliega("El número de probabilidades de muestra y observadas debe ser iguales")
ENDIF
Fin_Rutina_Chi2_Binomial

```

#### Rutina\_Chi2\_Discreta

```

Entra_Arreglo(Inicial, Final, Arreglo_Prob)
Entra_Arreglo(Inicial, Final, Arreglo_Observadas)
Sumatoria(Arreglo_Observadas, Ene)
Cuenta(Arreglo_Prob, Cuantas_Prob)
k = Cuantas_Prob
Chi2_Acumulada = 0
FOR I = 1 TO Cuantas_Prob DO
    Esperada[1] = Ene * Arreglo_Prob[1]
    Chi2[1] = (Arreglo_Observadas[1] - Esperada[1])2 / Esperada[1]
    Chi2_Acumulada = Chi2_Acumulada + Chi2[1]
ENDFOR
Grados_libertad = k - 1
Chi2_Cero = Obten_Chi2(Grados_Libertad, 0.5)
Despliega("El valor de  $\chi^2$  es" Chi2_Acumulada)
Despliega("El valor de  $\chi_0^2$  es" Chi2_Cero)
IF Chi2_Acumulada ≤ Chi2_Cero THEN
    Despliega(" $\chi^2 \leq \chi_0^2$ ")
    Despliega("Por lo tanto se acepta  $H_0$ ")
ELSE
    Despliega(" $\chi^2 > \chi_0^2$ ")
    Despliega("Por lo tanto se rechaza  $H_0$ ")

```

ENDIF  
Fin\_Rutina\_Chi2\_Discreta

Rutina\_Chi2\_Normal

```
Entra_Arreglo(Inicial,Final,Muestra)
Entra_Arreglo(Inicial,Final,Eventos_X_Muestra)
Cuenta(Muestra,Cuantos_Elementos)
Cuenta(Eventos_X_Muestra,Cuantos_EXM)
Sumatoria(Eventos_X_Muestra,Sumatoria_EXM)
IF Cuantos_Elementos = Cuantos_EXM THEN
  Contador = 1
  FOR I = 1 TO Cuantos_EXM DO
    FOR J = 1 TO Eventos_X_Muestra[I] DO
      Muestra_Completa[Contador] = Muestra[I]
      Contador = Contador+1
    ENDFOR
  ENDFOR
  Ene = Sumatoria_EXM
  Obten_Media(Muestra_Completa,Media)
  Obten_Variancia(Muestra_Completa,Variancia)
  Desviacion_Estandar =  $\sqrt{\text{Variancia}}$ 
  Chi2_Acumulada = 0
  FOR I = 1 TO Cuantos_Elementos DO
    Zeta[I] = (Muestra[I] - Media)/Desviacion_Estandar
    Normal[I]=Obten_Normal(Zeta[I],0.05)
    Esperada[I] = Ene*Normal[I]
  ENDFOR
  Indice_CFM = Cuantos_Elementos
  WHILE Indice_CFM $\neq$ 0 DO
    IF Esperada[Indice_CFM]<5 THEN
      Acumulada_CFM = 0
      Acumulada_O_CFM = 0
      WHILE Acum_CFM < 5 OR Indice_CFM  $\geq$  0 DO
        Acumulada_CFM = Acumulada_CFM + Esperada[I]
        Acumulada_O_CFM = Acumulada_O_CFM + Eventos_X_Muestra
        Esperada[Indice_CFM] = Acumulada_CFM
        Eventos_X_Muestra[Indice_CFM] = Acumulada_O_CFM
        Indice_CFM = Indice_CFM - 1
      ENDWHILE
    ELSE
      Indice_CFM = Indice_CFM - 1
    ENDIF
  ENDWHILE
  Chi2_Acumulada = 0
  FOR I = 1 TO Cuantos_Periodos DO
```

```

IF Esperanza[1] > 5 THEN
    Chi2[1] = (Eventos_X_Muestra[1] - Esperada[1])2/Esperada[1]
    Chi2_Acumulada = Chi2_Acumulada + Chi2[1]
    k = 1
ENDIF
ENDFOR
Grados_libertad = Cuantos_Elementos - 3
Chi2_Cero = Obten_Chi2(Grados_Libertad, 0.05)
Despliega("El valor de  $\chi^2$  es" Chi2_Acumulada)
Despliega("El valor de  $\chi_0^2$  es" Chi2_Cero)
IF Chi2_Acumulada ≤ Chi2_Cero THEN
    Despliega(" $\chi^2 \leq \chi_0^2$ ")
    Despliega("Por lo tanto se acepta  $H_0$ ")
ELSE
    Despliega(" $\chi^2 > \chi_0^2$ ")
    Despliega("Por lo tanto se rechaza  $H_0$ ")
ENDIF
ELSE
    Despliega("El número de probabilidades de muestra y observadas debe ser igual")
ENDIF
Fin_Rutina_Chi2_Normal

```

#### Rutina\_Chi2\_Poisson

```

Entra_Arreglo(Inicial, Final, Eventos_X_Periodo)
Entra_Arreglo(Inicial, Final, Periodos)
Cuenta(Eventos_X_Periodo, Cuantos_EXP)
Cuenta(Periodos, Cuantos_Periodos)
Sumatoria(Periodos, Sumatoria_Periodos)
Sumatoria_EXP = 0
k = Cuantos_EXP
FOR I = 1 TO Cuantos_Periodos DO
    Sumatoria_EXP = Sumatoria_EXP + (Eventos_X_Periodo[I] * Periodos[I])
ENDFOR
IF Cuantos_EXP = Cuantos_Periodos THEN
    Lambda = Sumatoria_EXP / Sumatoria_Periodos
    Poisson[0] = 0
    FOR I = 1 TO Cuantos_Periodos DO
        Poisson[I] = Obten_Poisson(Eventos_X_Periodo[I], Lambda)
        Poisson_SS[I] = Poisson[I] - Poisson[I-1]
        Esperada[I] = Sumatoria_Periodos * Poisson_SS[I]
    ENDFOR
    Indice_CFM = Cuantos_Periodos
    WHILE Indice_CFM ≠ 0 DO
        IF Esperada[Indice_CFM] < 5 THEN

```

```

Acumulada_CFM = 0
Acumulada_O_CFM = 0
WHILE Acum_CFM < 5 OR Indice_CFM ≥ 0 DO
    Acumulada_CFM = Acumulada_CFM + Esperada[1]
    Acumulada_O_CFM = Acumulada_O_CFM + Periodos[1]
    Esperada[Indice_CFM] = Acumulada_CFM
    Periodos[Indice_CFM] = Acumulada_O_CFM
    Indice_CFM = Indice_CFM - 1
ENDWHILE
ELSE
    Indice_CFM = Indice_CFM-1
ENDIF
ENDWHILE
Chi2_Acumulada = 0
FOR I = 1 TO Cuantos_Periodos DO
    IF Esperada[1] > 5 THEN
        Chi2[I] = (Periodos[1] - Esperada[1])2/Esperada[1]
        Chi2_Acumulada = Chi2_Acumulada+Chi2[1]
        k = I
    ENDIF
ENDFOR
Grados_libertad = k-2
Chi2_Cero = Obten_Chi2(Grados_Libertad,0.05)
Despliega("El valor de  $\chi^2$  es" Chi2_Acumulada)
Despliega("El valor de  $\chi_0^2$  es" Chi2_Cero)
IF Chi2_Acumulada ≤ Chi2_Cero THEN
    Despliega(" $\chi^2 \leq \chi_0^2$ ")
    Despliega("Por lo tanto se acepta  $H_0$ ")
ELSE
    Despliega(" $\chi^2 > \chi_0^2$ ")
    Despliega("Por lo tanto se rechaza  $H_0$ ")
ENDIF
ELSE
    Despliega("El número de probabilidades de muestra y observadas debe ser iguales")
ENDIF
Fin_Rutina_Chi2_Poisson

```

### Prueba de Hipótesis Medias

Rutina\_Prueba\_Medias\_1\_Muestra\_DE\_Conocida

Entra(Media\_Poblacional)

Entra\_Arreglo(Inicial,Final,Muestra)

Entra(Bilateral)

IF Bilateral = False THEN

```

    Entra(Media_Poblacional_Menor)
ENDIF
Cuenta(Muestra, Ene)
Obten_Media(Muestra, Media)
Obten_Variancia(Muestra, Variancia)
Desviacion_Estandar =  $\sqrt{\text{Variancia}}$ 
Zeta =  $((\text{Media\_Poblacional} - \text{Media}) * \sqrt{\text{Ene}}) / \text{Desviacion\_Estandar}$ 
Zeta_Bilateral = 1.96
Zeta_Unilateral = 1.645
IF Bilateral = True THEN
    Despliega("El valor de Z es" Zeta "y el valor de  $Z_{1,2}$  es"  $\pm$ Zeta_Bilateral)
    IF -Zeta_Bilateral  $\leq$  Zeta AND Zeta  $\leq$  Zeta_Bilateral THEN
        Despliega(Zeta "est dentro del intervalo  $Z_1 \leq Z \leq Z_2$ ")
        Despliega("Por lo tanto se acepta  $H_0$ ")
    ELSE
        Despliega(Zeta "no est dentro del intervalo  $Z_1 \leq Z \leq Z_2$ ")
        Despliega("Por lo tanto se rechaza  $H_0$ ")
    ENDIF
ELSE
    IF Media_Poblacional_Menor = True THEN
        Despliega("El valor de Z es" Zeta "y el valor de  $Z_2$  es" Zeta_Unilateral)
        IF Zeta  $\leq$  Zeta_Unilateral THEN
            Despliega("Entonces" Z  $\leq$   $Z_2$  ")
            Despliega("Por lo tanto se acepta  $H_0$ ")
        ELSE
            Despliega("Entonces" Z >  $Z_2$  ")
            Despliega("Por lo tanto se rechaza  $H_0$ ")
        ENDIF
    ELSE
        Despliega("El valor de Z es" Zeta "y el valor de  $Z_2$  es" -Zeta_Unilateral)
        IF -Zeta_Unilateral  $\leq$  Z THEN
            Despliega("Entonces" Z  $\leq$   $Z_1$  ")
            Despliega("Por lo tanto se acepta  $H_0$ ")
        ELSE
            Despliega("Entonces  $Z_1 > Z$ ")
            Despliega("Por lo tanto se rechaza  $H_0$ ")
        ENDIF
    ENDIF
ENDIF
Fin_Rutina_Prueba_Medias_1_Muestra_DE_Conocida

Rutina_Prueba_Medias_1_Muestra_DE_Desconocida
    Entra(Media_Poblacional)
    Entra_Arreglo(Inicial, Final, Muestra)

```

```

Entra(Bilateral)
IF Bilateral = False THEN
    Entra(Media_Poblacional_Menor)
ENDIF
Cuenta(Muestra, Ene)
Obten_Media(Muestra, Media)
Obten_Variancia(Muestra, Variancia)
Desviacion_Estandar =  $\sqrt{\text{Variancia}}$ 
Te = ((Media - Media_Poblacional) *  $\sqrt{\text{Ene}}$ ) / Desviacion_Estandar
Grados_Libertad = Ene - 1
t = Obten_t(Grados_Libertad, 0.05)
IF Bilateral = True THEN
    Despliega("El valor de t es" Te "y el valor de  $t_{1,2}$  es "  $\pm t$ )
    IF  $-t \leq \text{Te} \leq t$  THEN
        Despliega("Entonces  $t_1 \leq t \leq t_2$  ")
        Despliega("Por lo tanto se acepta  $H_0$  ")
    ELSE
        Despliega(Te "no está dentro del intervalo  $t_1 \leq t \leq t_2$  ")
        Despliega("Por lo tanto se rechaza  $H_0$  ")
    ENDIF
ELSE
    IF Media_Poblacional_Menor = True THEN
        Despliega("El valor de t es" Te "y el valor de  $t_1$  es"  $-t$ )
        IF  $-t \leq \text{Te}$  THEN
            Despliega("Entonces  $t_1 \leq t$ ")
            Despliega("Por lo tanto se acepta  $H_0$  ")
        ELSE
            Despliega("Entonces  $t_1 > t$ ")
            Despliega("Por lo tanto se rechaza  $H_0$  ")
        ENDIF
    ELSE
        Despliega("El valor de t es" Te "y el valor de  $t_2$  es" t)
        IF  $\text{Te} \leq t$  THEN
            Despliega("Entonces  $t \leq t_2$  ")
            Despliega("Por lo tanto se acepta  $H_0$  ")
        ELSE
            Despliega("Entonces  $t > t_2$  ")
            Despliega("Por lo tanto se rechaza  $H_0$  ")
        ENDIF
    ENDIF
ENDIF
Fin_Rutina_Prueba_Medias_1_Muestra_DE_Desconocida

```

```

Rotina_Prueba_Medias_2_Muestras_Indep_DE_Conocida
  Entra_Arreglo(Inicial,Final,Muestra_1)
  Entra_Arreglo(Inicial,Final,Muestra_2)
  Entra(Bilateral)
  IF Bilateral = False THEN
    Entra(Media_1_Mayor)
    Entra(Diferencia_Medias)
  ENDIF
  Cuenta(Muestra_1,Ene_1)
  Cuenta(Muestra_2,Ene_2)
  Obten_Media(Muestra_1,Media_1)
  Obten_Variancia(Muestra_1,Variancia_1)
  Desviacion_Estandar =  $\sqrt{\text{Variancia\_1}}$ 
  Obten_Media(Muestra_2,Media_2)
  Obten_Variancia(Muestra_2,Variancia_2)
  Desviacion_Estandar =  $\sqrt{\text{Variancia\_2}}$ 
   $\sigma d2 = (\text{Variancia\_1}/\text{Ene\_1}) + (\text{Variancia\_2}/\text{Ene\_2})$ 
   $\sigma d = \sqrt{\sigma d2}$ 
  IF Bilateral = False THEN
    Zeta = ((Media_1 -- Media_2) - Diferencia_Medias) /  $\sigma d$ 
  ELSE
    Zeta = (Media_1 - Media_2) /  $\sigma d$ 
  ENDIF
  Zeta_Bilateral = 1.96
  Zeta_Unilateral = 1.645
  IF Bilateral = True THEN
    Despliega("El valor de Z es" Zeta "y el valor de  $Z_{1,2}$  es"  $\pm$ Zeta_Bilateral)
    IF -Zeta_Bilateral  $\leq$  Zeta AND Zeta  $\leq$  Zeta_Bilateral THEN
      Despliega(Zeta "está dentro del intervalo  $Z_1 \leq Z \leq Z_2$ ")
      Despliega("Por lo tanto se acepta  $H_0$ ")
    ELSE
      Despliega(Zeta "no está dentro del intervalo  $Z_1 \leq Z \leq Z_2$ ")
      Despliega("Por lo tanto se rechaza  $H_0$ ")
    ENDIF
  ELSE
    IF Media_1_Mayor = False THEN
      Despliega("El valor de Z es" Zeta "y el valor de  $Z_1$  es" -Zeta_Unilateral)
      IF -Zeta_Unilateral  $\leq$  Zeta THEN
        Despliega("Entonces"  $Z_1 \leq Z$ )
        Despliega("Por lo tanto se acepta  $H_0$ ")
      ELSE
        Despliega("Entonces"  $Z_1 > Z$ )
        Despliega("Por lo tanto se rechaza  $H_0$ ")
      ENDIF
    ENDIF
  ENDIF

```



```

    ENDIF
ELSE
    Despliega("El valor de Z es" Zeta "y el valor de Z2 es" Zeta_Unilateral)
    IF Zeta ≤ Zeta_Unilateral THEN
        Despliega("Entonces" Z ≤ Z2 ")
        Despliega("Por lo tanto se acepta H0 ")
    ELSE
        Despliega("Entonces Z > Z2 ")
        Despliega("Por lo tanto se rechaza H0 ")
    ENDIF
ENDIF
Fin_Rutina_Prueba_Medias_2_Muestras_Indep_DE_Conocida

```

```

Rutina_Prueba_Medias_2_Muestras_Indep_DE_Desconocida
    Entra_Arreglo(Inicial,Final,Muestra_1)
    Entra_Arreglo(Inicial,Final,Muestra_2)
    Cuenta(Muestra_1,Ene_1)
    Cuenta(Muestra_2,Ene_2)
    IF Ene_1 ≤ 30 AND Ene_2 ≤ 30 THEN
        Obten_Media(Muestra_1,Media_1)
        Obten_Variancia(Muestra_1,Variancia_1)
        Desviacion_Estandar_1 = √Variancia_1
        Obten_Media(Muestra_2,Media_2)
        Obten_Variancia(Muestra_2,Variancia_2)
        Desviacion_Estandar_2 = √Variancia_2
        Sd2_Numerador = (Variancia_1*(Ene_1-1))+((Variancia_2*(Ene_2-1))
        Sd2_Denominador = ((Ene_1-1)+(Ene_2-1))
        Sd2_Factor = ((Ene_1+Ene_2)/(Ene_1*Ene_2))
        Sd2 = (Sd2Numerador/Sd2_Denominador)*Sd2_Factor
        Sd = √Sd2
        Prueba_F
        IF Variancias_Iguales = True THEN
            Te = (Media_1 - Media_2)/Sd
            Grados_Libertad = Ene_1 + Ene_2 - 2
        ELSE
            Te_Denominador = (Variancia_1/Ene_1 + Variancia_2/Ene_2)
            Te = (Media_1 - Media_2)/√Te_Denominador
            GL_Numerador = (Variancia_1/Ene_1 + Variancia_2/Ene_2)2
            GL_Denom_1 = (Variancia_1/Ene_1)2 / (Ene_1+1)
            GL_Denom_2 = (Variancia_2/Ene_2)2 / (Ene_2+1)
            Grados_Libertad = GL_Numerador/(GL_Denom_1 + GL_Denom_2)
            Grados_Libertad = Grados_Libertad - 2
        ENDIF
    ENDIF

```

```

t = Obten_t(Grados_Libertad,0.05)
Despliega("El valor de t es" Te "y el valor de t1,2 es " ± t)
IF -t ≤ Te AND Te ≤ t THEN
    Despliega("Entonces t1 ≤ t ≤ t2")
    Despliega("Por lo tanto se acepta H0")
ELSE
    Despliega("Te "no está dentro del intervalo t1 ≤ t ≤ t2")
    Despliega("Por lo tanto se rechaza H0")
ENDIF
ELSE
IF Ene_1 > 30 THEN
    Despliega("El número de elementos de la muestra 1 es mayor a 30")
ENDIF
IF Ene_2 > 30 THEN
    Despliega("El número de elementos de la muestra 2 es mayor a 30")
ENDIF
Fin_Rutina_Prueba_Medias_2_Muestras_Indep_DE_Desconocida

```

#### Rutina\_Prueba\_Medias\_2\_Muestras\_Pareadns

```

Entra_Arreglo(Inicial,Final,Muestra_1)
Entra_Arreglo(Inicial,Final,Muestra_2)
Cuenta(Muestra,Ene_1)
Cuenta(Muestra,Ene_2)
Sumatoria_Diferencias = 0
Sumatoria_Diferencias2 = 0
IF Ene_1 = Ene_2 THEN
    FOR i = 1 TO Ene_1 DO
        Diferencia[i] = Muestra_1[i] - Muestra_2[i]
        Diferencia2[i] = Diferencia[i]2
        Sumatoria_Diferencias = Sumatoria_Diferencias + Diferencia[i]
        Sumatoria_Diferencias2 = Sumatoria_Diferencias2 + Diferencia2[i]
    ENDFOR
     $\bar{d} = \text{Sumatoria\_Diferencias} / \text{Ene\_1}$ 
     $Sd\_1 = \text{Sumatoria\_Diferencias} / (\text{Ene\_1} - 1)$ 
     $Sd\_2 = (\text{Ene\_1} * \bar{d}^2) / (\text{Ene\_1} - 1)$ 
     $Sd = \sqrt{Sd\_1 - Sd\_2}$ 
     $Te = (\text{Sumatoria\_Diferencias} * \sqrt{\text{Ene\_1}}) / Sd$ 
    Grados_Libertad = Ene_1 - 1
    t = Obten_t(Grados_Libertad,0.05)
    Despliega("El valor de t es" Te "y el valor de t1,2 es " ± t)
IF -t ≤ Te AND Te ≤ t THEN
    Despliega("Entonces t1 ≤ t ≤ t2")
    Despliega("Por lo tanto se acepta H0")

```

```

ELSE
  Despliega(Te "no está dentro del intervalo  $t_1 \leq t \leq t_2$ ")
  Despliega("Por lo tanto se rechaza  $H_0$ ")
ENDIF
ELSE
  Despliega("El numero de elementos de las muestras debe ser el mismo")
ENDIF
Fin_Rutina_Prueba_Medias_2_Muestras_Pareadas

```

### Prueba de Hipótesis Variancias

#### Rutina\_Prueba\_Variancias\_1\_Muestra

```

  Entra_Arreglo(Inicial,Final,Muestra)

```

```

  Entra(Variancia_Poblacional)

```

```

  Entra(Bilateral)

```

```

  IF Bilateral = False THEN

```

```

    Entra(Variancia_Poblacional_Menor)

```

```

  ENDIF

```

```

  Cuenta(Muestra, Ene)

```

```

  Sumatoria(Muestra, Sumatoria_Muestra)

```

```

  Obten_Variancia(Muestra, Variancia_Muestral)

```

```

  DE_Muestral =  $\sqrt{\text{Variancia\_Muestral}}$ 

```

```

  Grados_Libertad = Ene-1

```

```

  Chi2 = (Grados_Libertad*Variancia_Muestral)/Variancia_Poblacional

```

```

  Chi2_Izq = Obten_Chi2(Grados_Libertad,0.975)

```

```

  Chi2_Der = Obten_Chi2(Grados_Libertad,0.025)

```

```

  Chi2_Unilateral = Obten_Chi2(Grados_Libertad,0.05)

```

```

  IF Bilateral = True THEN

```

```

    Despliega("El valor de  $\chi^2$  es" Chi2)

```

```

    Despliega("El valor de  $\chi^2_1$  es" Chi2_Izq)

```

```

    Despliega("El valor de  $\chi^2_2$  es" Chi2_Der)

```

```

    IF Chi2_Izq  $\leq$  Chi2 AND Chi2  $\leq$  Chi2_Der THEN

```

```

      Despliega(" $\chi^2_1 \leq \chi^2 \leq \chi^2_2$ ")

```

```

      Despliega("Por lo tanto se acepta  $H_0$ ")

```

```

    ELSE

```

```

      Despliega(Chi2 "no está dentro del intervalo  $\chi^2_1 \leq \chi^2 \leq \chi^2_2$ ")

```

```

      Despliega("Por lo tanto se rechaza  $H_0$ ")

```

```

    ENDIF

```

```

  ELSE

```

```

    Despliega("El valor de  $\chi^2$  es" Chi2)

```

```

    Despliega("El valor de  $\chi^2_2$  es" Chi2_Unilateral)

```

```

    IF Variancia_Poblacional_Menor = True THEN

```

```

      IF Chi2_Unilateral  $\leq$  Chi2 THEN

```

```

        Despliega(" $\chi^2 \leq \chi^2_2$ ")

```

```

    Despliega("Por lo tanto se acepta  $H_0$ ")
ELSE
    Despliega("  $\chi^2 > \chi_2^2$ ")
    Despliega("Por lo tanto se rechaza  $H_0$ ")
ENDIF
ELSE
    IF Chi2 ≤ Chi2_Unilateral THEN
        Despliega("  $\chi_1^2 \leq \chi^2$ ")
        Despliega("Por lo tanto se acepta  $H_0$ ")
    ELSE
        Despliega("  $\chi_1^2 > \chi^2$ ")
        Despliega("Por lo tanto se rechaza  $H_0$ ")
    ENDIF
ENDIF
Fin_Rutina_Prueba_Variancias_1_Muestra

Rutina_Prueba_Variancias_2_Muestra
    Entra_Arreglo(Inicial,Final,Muestra_1)
    Entra_Arreglo(Inicial,Final,Muestra_2)
    Cuenta(Muestra_1,Ene_1)
    Cuenta(Muestra_2,Ene_2)
    Obten_Variancia(Muestra,Variancia_Muestra_1)
    Obten_Variancia(Muestra,Variancia_Muestra_2)
    Grados_Libertad_1 = Ene_1-1
    Grados_Libertad_2 = Ene_2-1
    Entra(Bilateral)
    IF Variancia_1 < Variancia_2 THEN
        Variancia_Tempo = Variancia_1
        Variancia_1 = Variancia_2
        Variancia_2 = Variancia_Tempo
        Grados_Libertad_Tempo = Grados_Libertad_1
        Grados_Libertad_1 = Grados_Libertad_2
        Grados_Libertad_2 = Grados_Libertad_Tempo
    ENDIF
    Efe = Variancia_1/Variancia_2
    Efe_Der = Obten_Efe(Grados_Libertad_1,Grados_Libertad_2,0.025)
    Efe_Izq = Obten_Efe(Grados_Libertad_2,Grados_Libertad_1,0.025)
    Efe_Izq = 1/Efe_Izq
    Efe_Unilateral = Obten_Efe(Grados_Libertad_1,Grados_Libertad_2,0.05)
    IF Bilateral = True THEN
        Despliega("El valor de  $F$  es" Efe)
        Despliega("El valor de  $F_1$  es" Efe_Izq)
        Despliega("El valor de  $F_2$  es" Efe_Der)
        IF Efe_Izq ≤ Efe AND Efe ≤ Efe_Der THEN

```

```

    Despliega("  $F_1 \leq F \leq F_2$  ")
    Despliega("Por lo tanto se acepta  $H_0$  ")
ELSE
    Despliega(Efe "no está dentro del intervalo")
    Despliega("  $F_1 \leq F \leq F_2$  ")
    Despliega("Por lo tanto se rechaza  $H_0$  ")
ENDIF
ELSE
    Despliega("El valor de  $F$  es" Efe)
    Despliega("El valor de  $F_1$  es 1")
    Despliega("El valor de  $F_2$  es" Efe_Unilateral)
    IF  $1 \leq Efe$  AND  $Efe \leq Efe\_Der$  THEN
        Despliega("  $1 \leq F \leq F_2$  ")
        Despliega("Por lo tanto se acepta  $H_0$  ")
    ELSE
        Despliega(Efe "no está dentro del intervalo")
        Despliega("  $1 \leq F \leq F_2$  ")
        Despliega("Por lo tanto se rechaza  $H_0$  ")
    ENDIF
ENDIF
Fin_Rutina_Prueba_Variaciones_Dos_Muestra

```

### Prueba de Hipótesis Proporciones

#### Rutina\_Prueba\_Proporciones\_Una\_Muestra

Entra(*Proporcion\_Historica*)

Entra(*Proporcion\_Muestral*)

Entra(*Ene*)

Accurate =  $Ene * Proporcion\_Muestral$

IF Accurate > 5 THEN

Entra(*Bilateral*)

IF *Bilateral* = False THEN

Entra(*Proporcion\_Historica\_Menor*)

ENDIF

$\sigma^2 = (Proporcion\_Historica * (1 - Proporcion\_Historica)) / Ene$

$\sigma = \sqrt{\sigma^2}$

Zeta =  $(Proporcion\_Muestral - Proporcion\_Historica) / \sigma$

Zeta\_Bilateral = 1.96

Zeta\_Unilateral = 1.645

IF *Bilateral* = True THEN

Despliega("El valor de  $Z$  es" Zeta "y el valor de  $Z_{1,2}$  es"  $\pm Zeta\_Bilateral$ )

IF  $-Zeta\_Bilateral \leq Zeta$  AND  $Zeta \leq Zeta\_Bilateral$  THEN

Despliega(Zeta "está dentro del intervalo  $Z_1 \leq Z \leq Z_2$  ")

```

    Despliega("Por lo tanto se acepta  $H_0$ ")
ELSE
    Despliega("Zeta "no está dentro del intervalo  $Z_1 \leq Z \leq Z_2$ ")
    Despliega("Por lo tanto se rechaza  $H_0$ ")
ENDIF
ELSE
    IF Proporción_Histórica_Menor = True THEN
        Despliega("El valor de Z es" Zeta "y el valor de  $Z_2$  es" Zeta_Unilateral)
        IF Zeta  $\leq$  Zeta_Unilateral THEN
            Despliega("Entonces"  $Z \leq Z_2$ ;)
            Despliega("Por lo tanto se acepta  $H_0$ ")
        ELSE
            Despliega("Entonces"  $Z > Z_2$ ;)
            Despliega("Por lo tanto se rechaza  $H_0$ ")
        ENDIF
    ELSE
        Despliega("El valor de Z es" Zeta "y el valor de  $Z_1$  es" -Zeta_Unilateral)
        IF -Zeta_Unilateral  $\leq$  Z THEN
            Despliega("Entonces"  $Z_1 \leq Z$ )
            Despliega("Por lo tanto se acepta  $H_0$ ")
        ELSE
            Despliega("Entonces  $Z_1 > Z$ ")
            Despliega("Por lo tanto se rechaza  $H_0$ ")
        ENDIF
    ENDIF
ELSE
    Despliega("El producto de  $n$  por la proporción histórica no puede ser menor o
    igual a 5")
ENDIF
ENDIF
Fin_Rutina_Prueba_Proporciones_Una_Muestra

Rutina_Prueba_Proporciones_Dos_Muestras
  Entra(Proporción_Histórica_1)
  Entra(Proporción_Muestral_1)
  Entra(Ene_1)
  Entra(Proporción_Histórica_2)
  Entra(Proporción_Muestral_2)
  Entra(Ene_2)
  Accurate_1 = Ene_1*Proporción_Muestral_1
  Accurate_2 = Ene_2*Proporción_Muestral_2
  IF Accurate_2 > 5 AND Accurate_1 > 5 THEN
    Bilateral = False
    Entra(Bilateral)
  
```

```

σp2_1 = (Proporcion_Historica_1*(1 - Proporcion_Historica_1))/Ene_1
σp2_2 = (Proporcion_Historica_2*(1 - Proporcion_Historica_2))/Ene_2
σp2 = σp2_1 + σp2_2
σp = √σp2
IF Bilateral = False THEN
    Entra(Proporcion_Historica_1_Menor)
    Zeta_Numerador_1 = (Proporcion_Muestral_1-Proporcion_Muestral_2)
    Zeta_Numerador_2 = (Proporcion_Historica_1-Proporcion_Historica_2)
    Zeta = (Zeta_Numerador_1 - Zeta_Numerador_2)/σp
ELSE
    Zeta = (Proporcion_Muestral_1-Proporcion_Muestral_2)/σp
ENDIF
Zeta_Bilateral = 1.96
Zeta_Unilateral = 1.645
IF Bilateral = True THEN
    Despliega("El valor de Z es" Zeta "y el valor de Z1,2 es" ±Zeta_Bilateral)
    IF -Zeta_Bilateral ≤ Zeta AND Zeta ≤ Zeta_Bilateral THEN
        Despliega(Zeta "está dentro del intervalo Z1 ≤ Z ≤ Z2")
        Despliega("Por lo tanto se acepta H0")
    ELSE
        Despliega(Zeta "no está dentro del intervalo Z1 ≤ Z ≤ Z2")
        Despliega("Por lo tanto se rechaza H0")
    ENDIF
ELSE
    IF Proporcion_Historica_1_Menor = True THEN
        Despliega("El valor de Z es" Zeta "y el valor de Z2 es" Zeta_Unilateral)
        IF Zeta ≤ Zeta_Unilateral THEN
            Despliega("Entonces" Z ≤ Z2 ")
            Despliega("Por lo tanto se acepta H0")
        ELSE
            Despliega("Entonces" Z > Z2 ")
            Despliega("Por lo tanto se rechaza H0")
        ENDIF
    ELSE
        Despliega("El valor de Z es" Zeta "y el valor de Z2 es" -Zeta_Unilateral)
        IF -Zeta_Unilateral ≤ Z THEN
            Despliega("Entonces" Z1 ≤ Z")
            Despliega("Por lo tanto se acepta H0")
        ELSE
            Despliega("Entonces Z1 > Z")
            Despliega("Por lo tanto se rechaza H0")
        ENDIF
    ENDIF
ENDIF

```

```

ENDIF
ELSE
    Despliega("Los productos de  $n1$  y  $n2$  por las proporciones historicas 1 y 2 deben
        ser mayores que 5")
ENDIF
Fin_Rutina_Prueba_Proporciones_Dos_Muestras

```

### **Diseño y calidad del software**

En el proceso de diseño de software, la calidad del diseño se logra mediante series de revisiones técnicas formales. Para poder evaluar la calidad de la representación del diseño, se deben establecer criterios para el buen diseño.

- Un diseño debe de exhibir una organización jerárquica que haga un inteligente uso del control entre los elementos del software.
- Un diseño debe de ser modular; esto es, el software debe de estar particionado lógicamente en elementos que ejecuten funciones específicas y subfunciones.
- Un diseño debe de poseer módulos que exhiban características funcionales independientes.
- Un diseño debe ser derivado utilizando un método repetitivo dirigido por la información obtenida durante el análisis de requerimientos del software.

El proceso de diseño en ingeniería de software debe fomentar el buen diseño por medio de la aplicación de principios fundamentales de diseño, una metodología sistemática, y una minuciosa revisión.

### **CODIFICACIÓN Y LENGUAJE DE PROGRAMACIÓN**

Todos los pasos de ingeniería de software que se han realizado hasta este punto están orientados hacia un objetivo final: el traducir las representaciones de software en una forma que pueda ser entendida por la computadora. Este es el paso de codificación, un proceso que transforma el diseño en sentencias de un lenguaje de programación.

La gran mayoría de las aplicaciones de software utilizan lenguajes de alto nivel como lo son Ada, COBOL, Pascal o C.

Considerado como un paso en el proceso de ingeniería de software, la codificación es una consecuencia natural del diseño. Sin embargo, las características del lenguaje de programación y el estilo de codificación pueden afectar profundamente la calidad y la facilidad de mantenimiento del software.



## **Microsoft Excel**

Microsoft Excel nació como una hoja de cálculo, pero se ha convertido en una poderosa herramienta para la ejecución de múltiples tareas administrativas y financieras. Inmersa en la plataforma Windows de Microsoft, Excel utiliza un amigable ambiente gráfico como interfaz con el usuario para brindar las funciones de hoja de cálculo, base de datos, procesador de palabras, gráficas y macros. Las funciones principales y más completas de Excel son, por naturaleza, la hoja de cálculo y las macros; las demás funciones mencionadas son facilidades brindadas por código de otros productos de Microsoft como lo son Access, Word y Power Point. Este sistema está programado en versión 4.

## **Macros**

Una macro, virtualmente en todas las aplicaciones para computadora, es simplemente un método de automatizar una serie de tareas en un sólo comando definido por el usuario. La palabra es el nombre corto de macroinstrucción y ha venido siendo utilizada desde los tiempos de las perforadoras de tarjetas.

De cierta forma todos los programas de computadora están organizados en conjuntos de macros. Todas las funciones que se encuentran detrás de los menús o de los iconos son básicamente macros internas escritas y estructuradas.

De la misma forma Microsoft Windows ha "ensamblado" estas macros de una forma lógica y accesible a través de menús para que cualquier usuario pueda combinar varias funciones y comandos en rutinas para poder realizar tareas más eficientes y rápidas.

A continuación se presenta una lista de las ventajas que las macros ofrecen, y como estas ventajas ayudan a adecuar las facilidades de Microsoft Excel a necesidades de programación específicas:

Las macros puede hacer en minutos lo que tomaría horas realizar en una hoja de trabajo estándar. Muchos negocios por ejemplo usan macros simples para crear plantillas en pronósticos estandarizados y análisis (Recuperan apropiadamente los datos de otras hojas de cálculo, registra apropiadamente las fechas y las fórmulas dentro de las nuevas hojas y formatea las planillas correctamente). Sin macros, se pueden pasar una hora o más creando los formatos antes de empezar a capturar información.

Algunas veces las macros no pueden hacer el trabajo tan sencillo como lo haría cualquier persona, pero ahorran dinero y tiempo porque liberan de hacer trabajo. Una sencilla macro puede automatizar una tarea completa, las macros que imprimen gráficas son un buen ejemplo; aunque la macro no acelera la ejecución del trabajo, puede cargar la hoja, hacer los cálculos, imprimirla y cerrarla, cargar otra y así continuar. Todo sin intervención humana.

Además es posible ejecutar algunas tareas que, solo algunas veces al año son necesarias. Por

ejemplo calculo de impuestos trimestrales, análisis semestral de inventarios, y hacer ajustes de auditoria anual. Estos casos no suceden frecuentemente, y es fácil olvidar cual fue el ultimo procedimiento. El proceso de desarrollo de un sistema de macros para automatizar una tarea ofrece facilidades de documentarla, explicando que pasos se deben seguir, y porque esta tarea es necesaria, por que un dato es necesario, de donde proviene y hacia donde debe ir. También es posible crear y utilizar una interfaz de usuario para un sistema de macros, presentando pantallas de ayuda, menús, mensajes, y ventanas de dialogo para proporcionar y presentar información más fácilmente.

Es común cometer errores como lo son la obtención de datos de una columna equivocada, el copiado de estos en un mal lugar u olvidar como obtener una cifra. Una macro no comete este tipo de errores. Además es posible que la macro haga labores de verificación de información antes de convertirla parte de una hoja de trabajo. Por supuesto macros, puede generar errores, pero esos errores son, generalmente, fáciles de rastrear y arreglar.

Por último, es posible que una macro contenga a su vez macros. Y, a su vez es posible asignar una macro a una única tecla.

Así como hay grandes ventajas para programar en macros, existen algunas desventajas que es conveniente mencionar.

Las macros no son una panacea para los problemas de una hoja de cálculo:

Toma tiempo crear un macro, así como para determinar si es necesario del todo elaborar una. Puede tomar una considerable cantidad de horas el investigar, leer y practicar acerca de este tema. Adicionalmente aun cuando una macro trabaja perfectamente el trabajo está a la mitad, es necesario documentarla.

Son escasas las macros que no requieren mantenimiento ni actualizaciones. Por lo que es necesario modificarlas al paso del tiempo.

A diferencia de las hojas de cálculo, las macros necesitan ser escritas en hojas especiales, esto implica que se deben de manejar dos hojas en vez de una.

En Excel de Microsoft existen dos tipos de macros: comandos y funciones. Las macros de comandos son usualmente iniciadas para simular acciones, comandos, que seguiría cualquier usuario. Por ejemplo, seleccionar, cortar pegar, cambiar formatos, cargar y guardar hojas, imprimir, salvar a disco, etc. Las macros de funciones utilizan el conjunto de funciones estándar de Excel que, ya programadas propiamente se vuelven mucho más poderosas. El conjunto de funciones de macros de Excel incluyen las siguientes categorías: ingeniería, finanzas, fecha y tiempo, información, matemáticas y trigonometría, estadística, búsqueda y referencia, base de datos, texto, lógicas y DDE.

El conjunto de ambos tipos de macros nos permite definir tres tipos de macro de acuerdo a su aplicación.

Macros utilitarias: Son macros que son útiles en el día a día y que se requiere que estén disponibles todo el tiempo, sin importar el tipo de hoja de cálculo en la que se trabaje.

Macros de tarea específica: Estas son macros que ayudan en la ejecución de tareas que se requieren ocasionalmente y en ciertas hojas de cálculo.

Macros de aplicación completa: Son macros que cuando se combinan con menús, ventanas de diálogo y barras de herramientas, crean una aplicación por sí misma.

### Codificación

Para la codificación se utilizó una plantilla que "ordenó" la programación por columnas. Esta plantilla se muestra en la Figura 4.1.

Nombre	Macros	Comentarios
--------	--------	-------------

Figura 4.1 Plantilla de Codificación

A continuación se muestra el detalle de algunas rutinas que forman parte del sistema, se sigue el orden de columnas determinado por la figura anterior. Por ser una programación intensa se limita a algunos módulos típicos. Adicionalmente, se ha "formateado" el código en macros de forma que pudiera ser incluido en papel tamaño carta.

#### Módulo FI Archivo FI.XLM

Nombre de la Macro	Macros	Comentarios
ContinuaPresentacion	<pre> FIUNAM =DIIALOG.BOX(MenuPresentacion) =RETURN() </pre>	<p>Nombre de la Macro</p> <p>Despliega ventana diálogo</p> <p>Termina macro</p>

#### Módulo Modelos Archivo Modelos.XLM

Nombre de la Macro	Macros	Comentarios
ContinuaModelos	<pre> Modelos =ECHO(FALSE) =SET.VALUE(ModelosMenuChoice,1) =DIIALOG.BOX(ModelosMenu) =IF(ContinuaModelos=FALSE,HALT()) =IF(ModelosMenuChoice=1) = ModelosBinomial() =ELSE.IF(ModelosMenuChoice=2) = ModelosPascal() =ELSE.IF(ModelosMenuChoice=3) = ModelosPoisson() =ELSE.IF(ModelosMenuChoice=4) = ModelosExponencial() =ELSE() = ModelosNormal() =END.IF() =RETURN() </pre>	<p>Nombre de la macro</p> <p>Inicializa variable ventana</p> <p>Despliega ventana diálogo</p> <p>Opciones</p> <p>Termina macro</p>

### ModelosPoisson

```
=SET.VALUE(MPoissonX,"")
=SET.VALUE(MPoissonMedia,"")
=SET.VALUE(MPoissonFDA,"")
=SET.VALUE(MPoissonFMP,"")
=SET.VALUE(MPoissonOK,"OK")
=SET.VALUE(MPoissonCancelar,"Cancelar")
=SET.VALUE(MPoissonTrigger1,214)
=SET.VALUE(MPoissonTrigger2,208)
=SET.VALUE(MPoissonTrigger3,208)
=SET.VALUE(MPoissonTrigger4,205)
=SET.VALUE(MPoissonTrigger5,205)
ContinuaMPoisson =DIALOG.BOX(ModelosPoissonMenu)
=IF(ContinuaMPoisson=FALSE,HALT())
=IF(OR(MPoissonX<0,MPoissonMedia<=0))
= ALERT("La media y X deben ser mayores o iguales a cero.",3)
= GOTO(ContinuaMPoisson)
=END.IF()
=SET.VALUE(MPoissonTrigger1,14)
=SET.VALUE(MPoissonTrigger2,8)
=SET.VALUE(MPoissonTrigger3,8)
=SET.VALUE(MPoissonTrigger4,5)
=SET.VALUE(MPoissonTrigger5,5)
=SET.VALUE(MPoissonFDA,ROUND(POISSON(MPoissonX,MPoissonMedia,TRUE),
=SET.VALUE(MPoissonFMP,ROUND(POISSON(MPoissonX,MPoissonMedia,FALSE)
=SET.VALUE(MPoissonOK,"Otro")
=SET.VALUE(MPoissonCancelar,"Terminar")
ContinuaMPoisson2 =DIALOG.BOX(ModelosPoissonMenu)
=IF(ContinuaMPoisson2=FALSE)
= HALT()
=ELSE()
= GOTO(ModelosPoisson)
=END.IF()
=RETURN()
```

### Módulo HipótesisMedias

### Archivo Medias.XLM

#### HMMuestrasPareadas

```
=SET.VALUE(HMMuestrasPareadasRangoMuestra_1,"")
=SET.VALUE(HMMuestrasPareadasRangoMuestra_2,"")
=SET.VALUE(HMMuestrasPareadasUnilateral,FALSE)
ContinuaHMMPareadas =DIALOG.BOX(HMMPareadasMenu)
=IF(ContinuaHMMPareadas=FALSE,HALT())
=IF(OR(HMMuestrasPareadasRangoMuestra_1=""HMMuestrasPareadasRangoMuestra_2=""
= ALERT("Se deben proporcionar argumentos de entrada válidos.
Utilice el icono otra vez.",3)
= HALT()
=END.IF()
=IF(HMMuestrasPareadasUnilateral=TRUE)
```

ContinualHMMParceadasUnilateral

```
= SET.VALUE(HMMuestrasParceadasMPMenor,1)
= DIALOG.HOX(HMMuestrasParceadasUnilateralMenu)
= IF(ContinualHMMParceadasUnilateral=FALSE,HALT())
=END.IF()
=FC10(FALSE)
=SELECT("R200C200:R350C250")
=CLEAR(1)
=DEFINE.NAME("HMMuestrasParceadasRangoMuestraCuantos_1","=R200C200")
=DEFINE.NAME("HMMuestrasParceadasRangoMuestraCuantos_2","=R201C200")
=SELECT(GET.CELL(5,HMMuestrasParceadasRangoMuestra_1))
=SET.VALUE(HMMuestrasParceadasRangoMuestraCuantos_1,COUNT(SELECTION()))
=CUT()
=SELECT("R200C201")
=PASTE()
=SELECT(GET.CELL(5,HMMuestrasParceadasRangoMuestra_2))
=SET.VALUE(HMMuestrasParceadasRangoMuestraCuantos_2,COUNT(SELECTION()))
=CUT()
=SELECT("R200C202")
=PASTE()
=SELECT("R1C1:R"&HMMuestrasParceadasRangoMuestraCuantos_1+14&"C25")
=CLEAR(1)
=SELECT("R1C2")
=COLUMN.WIDTH(22)
=PATTERNS(1,0,0)
=FORMULA("Prueba de Medias")
=FORMAT.FONT("MS Sans Serif",12,TRUE,TRUE,FALSE,FALSE,2)
=SELECT("R2C2")
=FORMULA("Dos Muestras Parceadas")
=FORMAT.FONT("MS Sans Serif",9,TRUE,TRUE,FALSE,FALSE,0)
=SELECT("R4C3")
=COLUMN.WIDTH(16)
=FORMULA("Ocurrencia")
=FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
=ALIGNMENT(3,FALSE,3,0)
=SELECT("R4C4")
=COLUMN.WIDTH(16)
=FORMULA("Muestra 1")
=FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
=ALIGNMENT(3,FALSE,3,0)
=SELECT("R4C5")
=COLUMN.WIDTH(16)
=FORMULA("Muestra 2")
=FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
=ALIGNMENT(3,FALSE,3,0)
=SELECT("R4C6")
=COLUMN.WIDTH(16)
=FORMULA("Unilateral")
=FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
=ALIGNMENT(3,FALSE,3,0)
=IF(HMMuestrasParceadasUnilateral=TRUE)
```

```

= SELECT("R7C6")
= FORMULA("Media 1")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R8C6")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= IF(HMMuestrasPareadasMPMejor=1)
= FORMULA("Menor")
= ELSE()
= FORMULA("Mayor")
= END.IF()
=END.IF()
=FOR("ContaOcorre",1,HMMuestrasPareadasRangoMuestraCuantos_1)
= SELECT("R"&4+ContaOcorre&"C3")
= FORMULA(ContaOcorre)
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
=NEXT()
=SELECT("R200C201:R"&200+
HMMuestrasPareadasRangoMuestraCuantos_1-1&"C201")
=FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
=CUT()
=SELECT("R5C4")
=PASTE()
=ALIGNMENT(3,FALSE,3,0)
=SELECT("R200C202:R"&200+
HMMuestrasPareadasRangoMuestraCuantos_2-1&"C202")
=FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
=CUT()
=SELECT("R5C5")
=PASTE()
=ALIGNMENT(3,FALSE,3,0)
=IF(HMMuestrasPareadasRangoMuestraCuantos_1 =
HMMuestrasPareadasRangoMuestraCuantos_2)
= DEFINE.NAME("HMMuestrasPareadasSumaDiferen","=R200C203")
= DEFINE.NAME("HMMuestrasPareadasSumaDiferenCuad","=R201C203")
= DEFINE.NAME("HMMuestrasPareadas1","=R202C203")
= DEFINE.NAME("HMMuestrasPareadasSd_1","=R203C203")
= DEFINE.NAME("HMMuestrasPareadasSd_2","=R204C203")
= DEFINE.NAME("HMMuestrasPareadasSd","=R205C203")
= DEFINE.NAME("HMMuestrasPareadasTe","=R206C203")
= DEFINE.NAME("HMMuestrasPareadasGl","=R207C203")
= DEFINE.NAME("HMMuestrasPareadas1","=R208C203")
= DEFINE.NAME("HMMuestrasPareadas_1","=R209C203")
= DEFINE.NAME("HMMuestrasPareadas_2","=R210C203")
= DEFINE.NAME("HMMuestrasPareadasDiferencia","=R211C203")
= DEFINE.NAME("HMMuestrasPareadasDiferenciaCuadrada","=R212C203")
= SET.VALUE(HMMuestrasPareadasSumaDiferen,0)
= SET.VALUE(HMMuestrasPareadasSumaDiferenCuad,0)
= SELECT("R4C7")

```

```

= COLUMN.WIDTH(16)
= FORMULA("Diferencia di")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R4C8")
= COLUMN.WIDTH(16)
= FORMULA("Diferencia di cuadrada")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= FOR("ContaOcorre",5,1,1MMuestrasPareadasRangoMuestraCuantos_1+4)
= SELECT("R"&ContaOcorre&"C4")
= SET.VALUE(1,1MMuestrasPareadas_1,GET.CELL(5,SELECTION()))
= SELECT("R"&ContaOcorre&"C5")
= SET.VALUE(1,1MMuestrasPareadas_2,GET.CELL(5,SELECTION()))
= SET.VALUE(1,1MMuestrasPareadasDiferencia,1MMuestrasPareadas_1-
1MMuestrasPareadas_2)
= SET.VALUE(1,1MMuestrasPareadasDiferenciaCuadrada,
SUMSQ(1MMuestrasPareadas_1-1MMuestrasPareadas_2))
= SET.VALUE(1,1MMuestrasPareadasSumaDiferen,
1MMuestrasPareadasSumaDiferen+1MMuestrasPareadasDiferencia)
= SET.VALUE(1,1MMuestrasPareadasSumaDiferenCuad,
1MMuestrasPareadasSumaDiferenCuad+
1MMuestrasPareadasDiferenciaCuadrada)
= SELECT("R"&ContaOcorre&"C7")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= FORMULA(1MMuestrasPareadasDiferencia)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R"&ContaOcorre&"C8")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= FORMULA(1MMuestrasPareadasDiferenciaCuadrada)
= ALIGNMENT(3,FALSE,3,0)
= NEXT()
= SELECT("R"&ContaOcorre&"C7")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= BORDER(0,0,0,1,0,,,,,0)
= FORMULA(1MMuestrasPareadasSumaDiferen)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R"&ContaOcorre&"C8")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= BORDER(0,0,0,1,0,,,,,0)
= FORMULA(1MMuestrasPareadasSumaDiferenCuad)
= ALIGNMENT(3,FALSE,3,0)
= SET.VALUE(1,1MMuestrasPareadasd,
1MMuestrasPareadasSumaDiferen/
1MMuestrasPareadasRangoMuestraCuantos_1)
= SET.VALUE(1,1MMuestrasPareadasSd_1,
1MMuestrasPareadasSumaDiferenCuad/
(1MMuestrasPareadasRangoMuestraCuantos_1-1))
= SET.VALUE(1,1MMuestrasPareadasSd_2,
1MMuestrasPareadasRangoMuestraCuantos_1*SUMSQ(1MMuestrasPareadasd))
= SET.VALUE(1,1MMuestrasPareadasSd_2,
1MMuestrasPareadasSd_2/(1MMuestrasPareadasRangoMuestraCuantos_1-1))

```

```

= SET.VALUE(H1MMuestrasPareadasSd,
SQRT(H1MMuestrasPareadasSd_1-H1MMuestrasPareadasSd_2))
= SET.VALUE(H1MMuestrasPareadasTe,
H1MMuestrasPareadasd*SQRT(H1MMuestrasPareadasRangoMuestraCuantos_1))
= SET.VALUE(H1MMuestrasPareadasTe,
H1MMuestrasPareadasTe/H1MMuestrasPareadasSd)
= SET.VALUE(H1MMuestrasPareadasGl,
H1MMuestrasPareadasRangoMuestraCuantos_1-1)
= SELECT("R5C6")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= IF(H1MMuestrasPareadasUnilateral= TRUE)
= FORMULA("Si")
= SET.VALUE(H1MMuestrasPareadasT,TINV(0.1,H1MMuestrasPareadasGl))
= ELSE()
= FORMULA("No")
= SET.VALUE(H1MMuestrasPareadasT,TINV(0.05,H1MMuestrasPareadasGl))
= END.IF()
= SELECT("R4C9")
= FORMULA("Media d:")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R4C10")
= FORMULA(ROUND(H1MMuestrasPareadasd,4))
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R5C9")
= FORMULA("Desviación Estándar Sd:")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R5C10")
= FORMULA(ROUND(H1MMuestrasPareadasSd,4))
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R6C9")
= FORMULA("Nivel de Significancia:")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R6C10")
= FORMULA("0.5")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R7C9")
= FORMULA("Grados de Libertad:")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R7C10")
= FORMULA(H1MMuestrasPareadasGl)
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,FALSE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R4C13")

```



```

= FORMULA("I_Cero: delta = 0")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= IF(IJMuestrasParedasUnilateral=FALSE)
= SELECT("R4C11")
= COLUMN.WIDTH(16)
= FORMULA("El valor de t es:")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R5C11")
= FORMULA("El valor de t_Uno es:")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R6C11")
= FORMULA("El valor de t_Dos es:")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R4C12")
= COLUMN.WIDTH(16)
= FORMULA(ROUND(IJMuestrasParedasTe,4))
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R5C12")
= FORMULA("·"&ROUND(HMMuestrasParedast,4))
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R6C12")
= FORMULA("·"&ROUND(IJMuestrasParedast,4))
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R5C13")
= FORMULA("I_Uno: delta < 0")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= IF(AND(-IJMuestrasParedast<=IHMuestrasParedasTe,
        IJMuestrasParedasTe<=IHMuestrasParedast))
= SELECT("R6C12")
= COLUMN.WIDTH(41)
= FORMULA("El valor de t está dentro del intervalo t_Uno <= t <= t_Dos")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R7C12")
= FORMULA("Por lo tanto se acepta I_Cero")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= ELSE()
= SELECT("R6C13")
= COLUMN.WIDTH(50)
= FORMULA("El valor de t no está dentro del intervalo t_Uno <= t <= t_Dos")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)

```

```

= ALIGNMENT(2,FALSE,3,0)
= SELECT("R7C13")
= FORMULA("Por lo tanto se rechaza H_Cero")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= END.IF()
= ELSE()
= SELECT("R4C11")
= COLUMN.WIDTH(16)
= FORMULA("El valor de t es")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R4C12")
= COLUMN.WIDTH(16)
= FORMULA(ROUND(HMMuestrasPareadasTe,4))
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= IF(HMMuestrasPareadasMPMenor=1)
= SELECT("R5C11")
= COLUMN.WIDTH(16)
= FORMULA("El valor de t_Dos es")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R5C12")
= FORMULA("+"&ROUND(HMMuestrasPareadast,4))
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R5C13")
= FORMULA("H_Uno: delta < 0")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= IF(HMMuestrasPareadasTe<= HMMuestrasPareadast)
= SELECT("R6C13")
= COLUMN.WIDTH(25)
= FORMULA("Puesto que t <= t_Dos")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R7C13")
= FORMULA("Por lo tanto se acepta H_Cero")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= ELSE()
= SELECT("R6C13")
= COLUMN.WIDTH(25)
= FORMULA("Puesto que t > t_Dos")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R7C13")
= FORMULA("Por lo tanto se rechaza H_Cero")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)

```

```

= ALIGNMENT(2,FALSE,3,0)
= END.IF()
= ELSE()
= SELECT("R5C11")
= COLUMN.WIDTH(16)
= FORMULA("E) valor de t_Uno es")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R5C12")
= COLUMN.WIDTH(16)
= FORMULA("=-&ROUND(X(11MMuestrasPareadas,4))")
= FORMAT.FONT("MS Sans Serif",8,FALSE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(3,FALSE,3,0)
= SELECT("R5C13")
= FORMULA("11_Uno: delta > 0")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= IF(-11MMuestrasPareadas<= 11MMuestrasPareadasTe)
= SELECT("R6C13")
= COLUMN.WIDTH(25)
= FORMULA("Puesto que t_Uno <= 1")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R7C13")
= FORMULA("Por lo tanto se acepta H_ Cero")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= ELSE()
= SELECT("R6C13")
= COLUMN.WIDTH(25)
= FORMULA("Puesto que t_Uno > 1")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= SELECT("R7C13")
= FORMULA("Por lo tanto se rechaza H_ Cero")
= FORMAT.FONT("MS Sans Serif",8,TRUE,FALSE,TRUE,FALSE,0)
= ALIGNMENT(2,FALSE,3,0)
= END.IF()
= END.IF()
= END.IF()
=ELSE()
= SELECT("R4C6")
= COLUMN.WIDTH(36)
= FORMULA("El tamaño de las muestras debe ser igual!")
= FORMAT.FONT("MS Sans Serif",10,FALSE,FALSE,TRUE,FALSE,0)
=END.IF()
=DELETE.NAME("11MMuestrasPareadasRangoMuestraCuantos_1")
=DELETE.NAME("11MMuestrasPareadasRangoMuestraCuantos_2")
=DELETE.NAME("11MMuestrasPareadasd")
=DELETE.NAME("11MMuestrasPareadasSd_1")

```

```

=DELETE.NAME("HMMuestrasPareadasSd_2")
=DELETE.NAME("HMMuestrasPareadasSd")
=DELETE.NAME("HMMuestrasPareadasI")
=DELETE.NAME("HMMuestrasPareadasIc")
=DELETE.NAME("HMMuestrasPareadasGd.")
=DELETE.NAME("HMMuestrasPareadasSumatDiferen")
=DELETE.NAME("HMMuestrasPareadasSumatDiferenCuad")
=DELETE.NAME("HMMuestrasPareadas_1")
=DELETE.NAME("HMMuestrasPareadas_2")
=DELETE.NAME("HMMuestrasPareadasDiferencia")
=DELETE.NAME("HMMuestrasPareadasDiferenciaCuadrada")
=DELETE.NAME("ContaOcurre")
=SELECT("RIC2")
=RETURN()

```

## PRUEBAS

La ejecución de pruebas es una tarea interesante para el ingeniero de software. Durante las fases de definición y desarrollo, el ingeniero intenta construir software partiendo de un concepto abstracto hacia una implantación tangible. Lo siguiente es probar. El ingeniero crea una serie de casos de prueba cuyo objetivo es "demoler" el software que ha creado. De hecho, probar es el paso del proceso de la ingeniería de software que puede ser visto como destructivo, más no constructivo. Los desarrolladores de software son, por naturaleza, creadores. El proceso de prueba requiere que el desarrollador elimine aquellas nociones preconcebidas de "correctitud" acerca de su software.

### Objetivo de la prueba

Los objetivos del paso de prueba son:

1. La prueba es el proceso de ejecutar el programa con la intención de encontrar un error.
2. Un buen caso de prueba es aquel que tiene una alta probabilidad de encontrar un error todavía no descubierto.

Si la prueba es conducida exitosamente, se descubrirán errores en el software. Y, como un beneficio secundario, la prueba demostrará que el sistema funciona de acuerdo a la especificaciones y que cumple los requerimientos. Pero, existe un aspecto que la prueba no puede realizar:

La prueba no puede mostrar la ausencia de defectos; esta solo puede mostrar que los defectos en el software están presentes.

A continuación se incluyen las pruebas realizadas para cada uno de los módulos. Son acompañadas por ejemplos prácticos que demuestran su utilidad.

## Información FI UNAM

Al seleccionar, con el mouse, el icono fi se obtendrá la pantalla de la Figura 4.2. Todos los módulos del sistema se accesan mediante mouse e iconos.

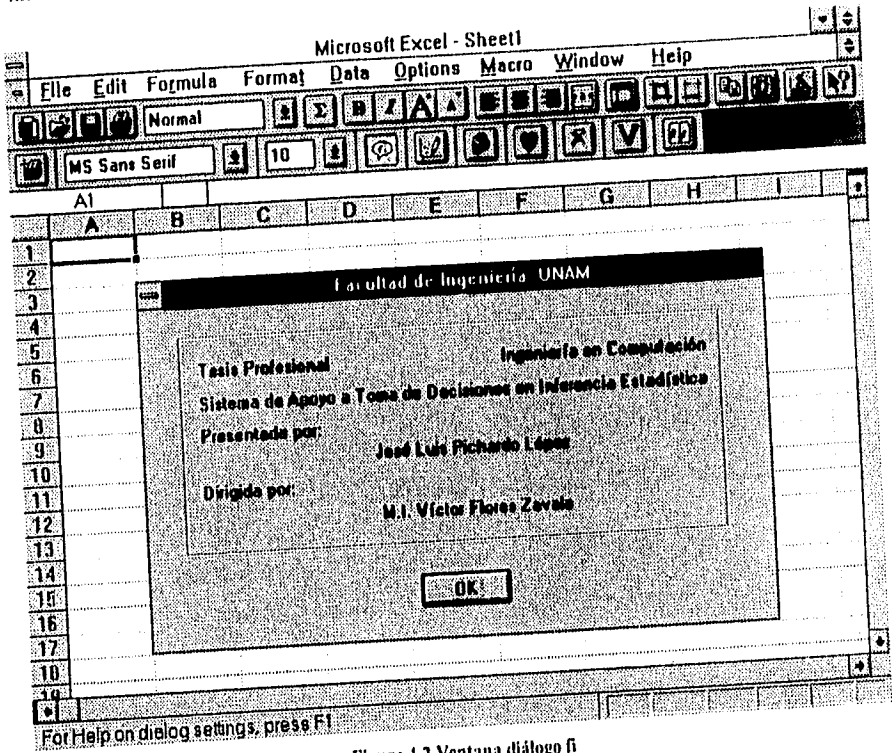


Figura 4.2 Ventana diálogo fi

## Parámetros Estadística Descriptiva

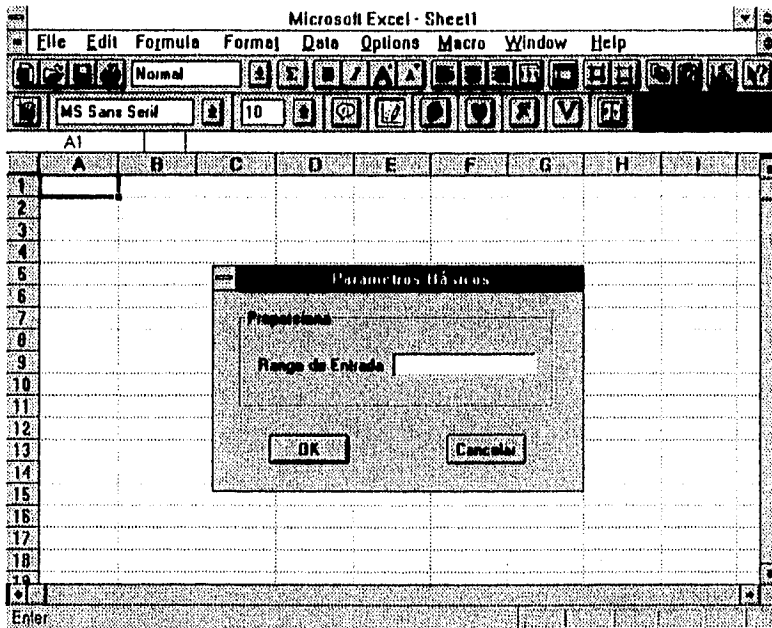


Figura 4.3 Ventana diálogo Parámetros Básicos

### Muestras de entrada:

Muestra 1		Muestra 2					
53	23	70	2	13	46	76	97
43	33	80	99	69	18	80	77
35	43	44	22	12	5	94	99
63	53	70	23	58	90	96	93
25	38	29	29	34	33	47	77
45		57	83	28	85	84	
54		42	79	27	85	48	
46		21	36	64	86	96	
39		72	40	58	97	76	
55		8	51	65	98	24	
51		73	54	77	96	49	
47		33	81	77	51	88	
47		70	1	16	77	84	
35		80	7	80	95	90	
29		59	99	70	62	90	

Las siguientes dos páginas son el resultado de las dos ejecuciones.

**Parámetros Básicos**  
Una muestra

<u>Muestra</u>	<u>Número de Elementos</u>	
53	<u>Número de Elementos</u>	20
43	<u>Mínimo</u>	23
35	<u>Máximo</u>	63
63	<u>Rango</u>	40
25	<u>Media</u>	42.85
45	<u>Mediana</u>	44
54	<u>Moda</u>	53
46	<u>Variación</u>	113.818
39	<u>Desviación Estándar</u>	10.6686
55	<u>Coef. Pearson</u>	0
51	<u>Básicos Curtosis</u>	-0.5291
47	<u>Básicos Asimetría</u>	-0.1966
47		
35		
29		
23		
33		
43		
53		
38		

Tabla de Frecuencias

<u>Clase</u>	<u>Frecuencia</u>	<u>%</u>	<u>Acumulada</u>	<u>% Rel. Acum</u>
30	3	0.15	3	0.15
40	5	0.25	8	0.4
50	6	0.3	14	0.7
60	5	0.25	19	0.95
	1	0.05	20	1
<u>Total</u>	20			

**Parámetros Básicos**

Una muestra

Muestra			
70	13	<u>Número de Elementos</u>	60
80	69	<u>Mínimo</u>	1
44	12	<u>Máximo</u>	99
70	58	<u>Rango</u>	98
29	34	<u>Media</u>	54.7667
57	28	<u>Mediana</u>	58.5
42	27	<u>Moda</u>	70
21	64	<u>Variación</u>	845.1311
72	58	<u>Desviación Estándar</u>	29.0711
8	65	<u>Coef. Pearson</u>	0
73	77	<u>Básicos Curtosis</u>	-1.1384
33	77	<u>Básicos Asimetría</u>	-0.2565
70	16		
80	80		
59	70		
2	46		
99	13		
22	5		
23	90		
29	33		
83	85		
79	85		
36	86		
40	97		
51	98		
54	96		
81	51		
1	77		
7	95		
99	62		

**Tabla de Frecuencias**

<u>Clase</u>	<u>Frecuencia</u>	<u>%</u>	<u>Acumulada</u>	<u>% Rel Acum</u>
10	5	0.0833	5	0.0833
20	4	0.0667	9	0.15
30	7	0.1167	16	0.2667
40	5	0.0833	21	0.35
50	3	0.05	24	0.4
60	7	0.1167	31	0.5167
70	8	0.1333	39	0.65
80	9	0.15	48	0.8
90	6	0.1	54	0.9
	6	0.1	60	1
<u>Total</u>	60			



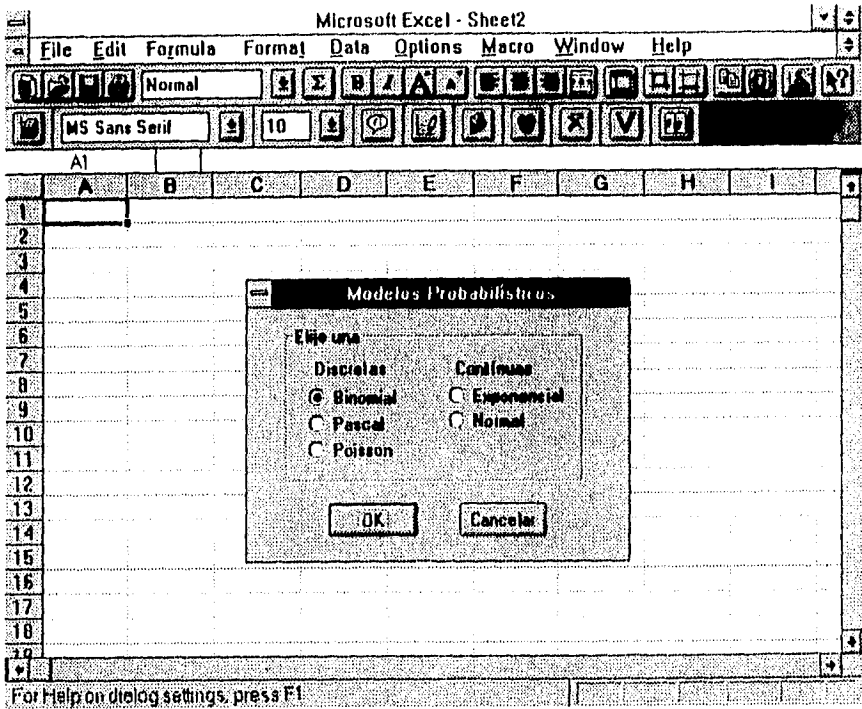


Figura 4.4 Ventana diálogo menú principal de Modelos Probabilísticos

## Binomial

La fábrica de ligas El Elefante S.A. produce ligas de hule con una alta resistencia a rompeduras. El Sr. Darío Gutiérrez, Jefe de Producción, ha encargado a el Departamento de Control de Calidad hacer una prueba tomando una muestra de 50 ligas de la línea de producción.

Se hicieron las pruebas de elasticidad y resultaron 3 ligas rotas.

La probabilidad  $p$  se estima:  $p = 3/50 = 0.06$

¿Cuál es la probabilidad de que más de 3 ligas se rompan en una muestra de 50?

Si se proporciona al sistema los datos del problema se obtendrán los resultados de la Figura 4.5.

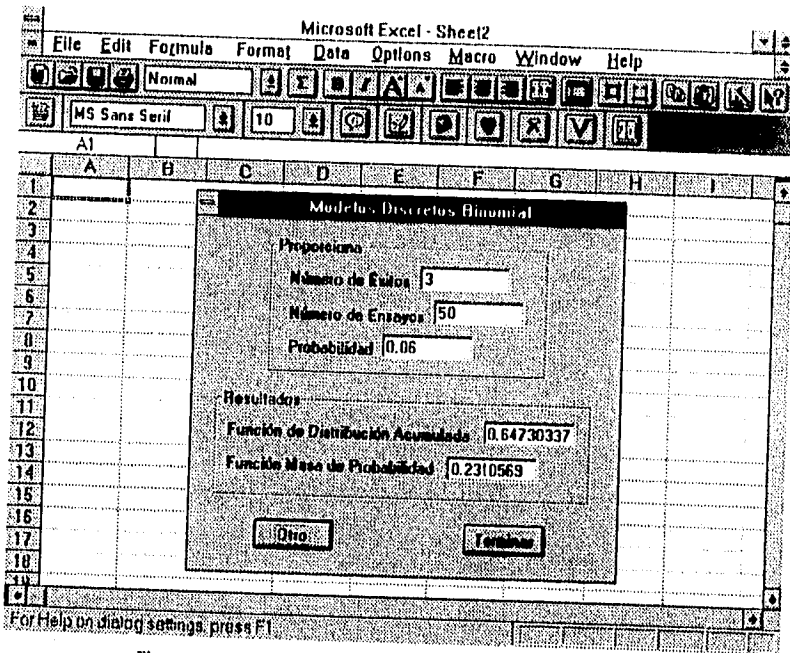


Figura 4.5 Resultado del problema de ligas de hule "El Elefante"

Por lo tanto la probabilidad de que más de 3 ligas se rompan es 0.2311.

## Exponencial

La Distribuidora de Pinturas Profesionales e Impermeabilizasteis S.A. de C.V. realiza un estudio para evaluar la posibilidad de contratar un empleado de mostrador extra. Por tener productos de alto consumo estacional, analizan optimizar el tiempo de atención de acuerdo a las llegadas de los clientes.

Según las observaciones del encargado llegan 3 personas por minuto. Por lo que se preguntan:

¿Cuál es la probabilidad de que el tiempo entre dos llegadas consecutivas sea menor que medio minuto?

Según la Figura 4.6 es la probabilidad de que esto suceda es de 0.7769.

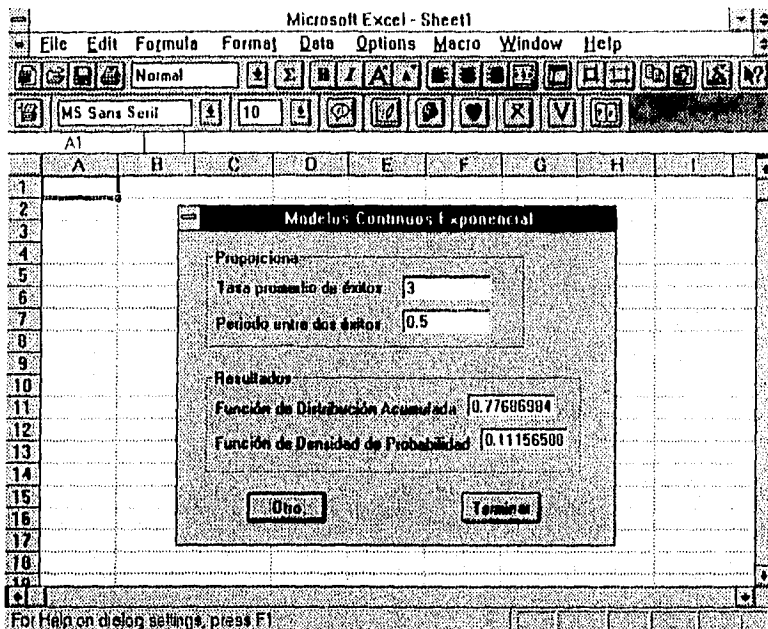


Figura 4.6 Modelos Probabilísticos Exponencial

### Normal

Aerovías de México tiene registrado que el tiempo necesario para que un viajero se traslade de México a Santiago de Chile es de 13.5 hrs. en promedio, con una variación de más o menos 1 hr.

Con la entrada de más compañías que brindan servicios de transportación aérea, el viajero ahora tiene más opciones para elegir pero generalmente son con conexiones en Miami.

Aeroméxico desea obtener una ventaja competitiva garantizando el tiempo de vuelo. Por lo que la Dirección Operaciones junto con la de Comercialización desean saber cuál es la probabilidad de que el recorrido tarde 14 hrs.

La probabilidad es 0.6915, de acuerdo a la Figura 4.7.

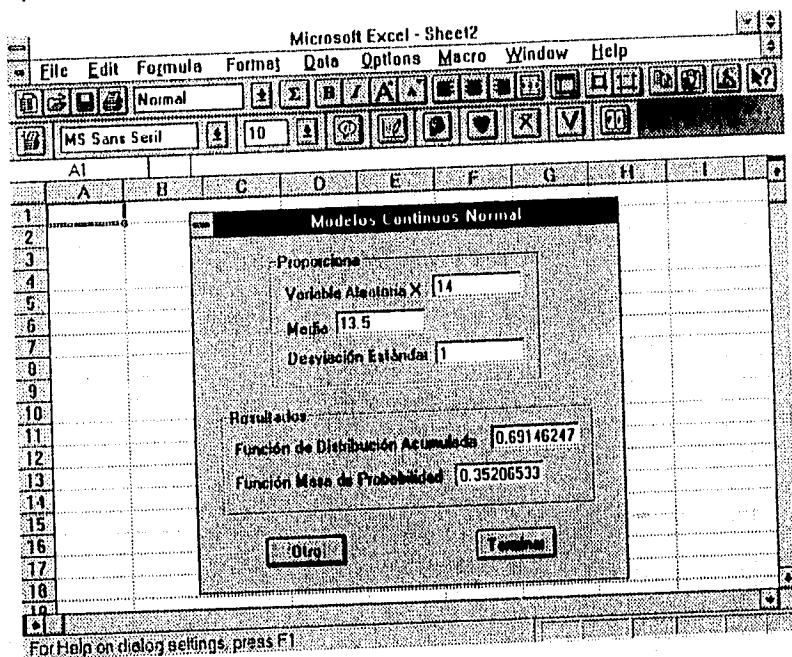


Figura 4.7 Modelos Probabilísticos Normal

## Pascal

Los estudios de cine Churubusco Azteca buscan, para conformar un grupo de "dobles" de actores, 10 individuos con excelentes reflejos, visión perfecta e inigualable condición física. Saben que la probabilidad de que un candidato esté altamente calificado es 0.25.

¿Cuál es la probabilidad de después de haber entrevistado a 5, los 10 siguientes cumplan los requisitos?

Si se proporciona al sistema los datos del problema se obtendrán los resultados de la Figura 4.8.

Por lo que la probabilidad es 0.5504866

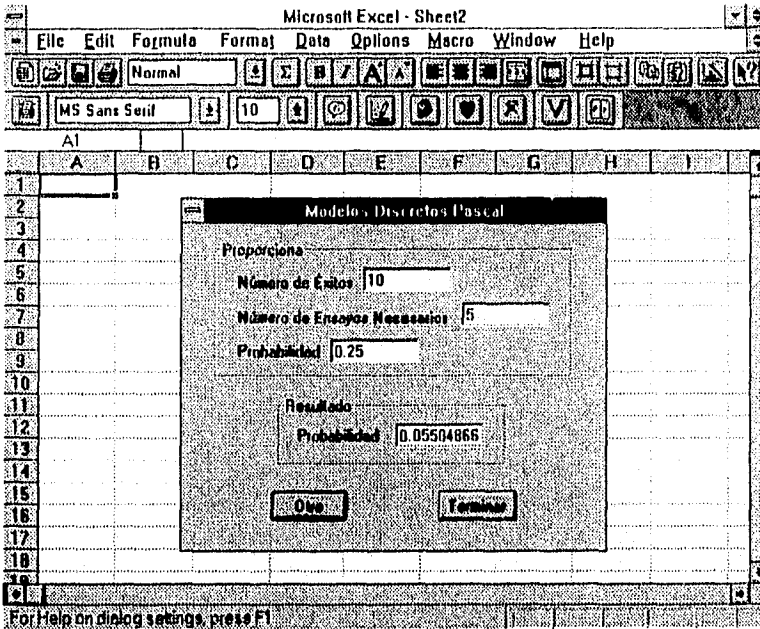


Figura 4.8 Modelos Probabilísticos Pascal

### Poisson

Federico y Gerardo son ingenieros que trabajan en el Colegio de Ingenieros Civiles A.C. Han determinado que una de las principales, y más costosas, causas de retraso en las construcciones son los accidentes. Basados en un estudio en las obras de ampliación del Metro, observaron que el número de accidentes serios que sufren los obreros de la construcción son de acuerdo a la siguiente tabla:

Accidentes por mes	0	1	2	3	4	5
Frecuencia	27	12	8	2	1	0

De aquí obtuvieron que el promedio de accidentes es  $38/50 = 0.76$ . Y se cuestionan:  
¿Cuál es la probabilidad de que suceda sólo un accidente serio al mes?

Según la figura 4.9 la probabilidad de que esto ocurra es 0.3554.

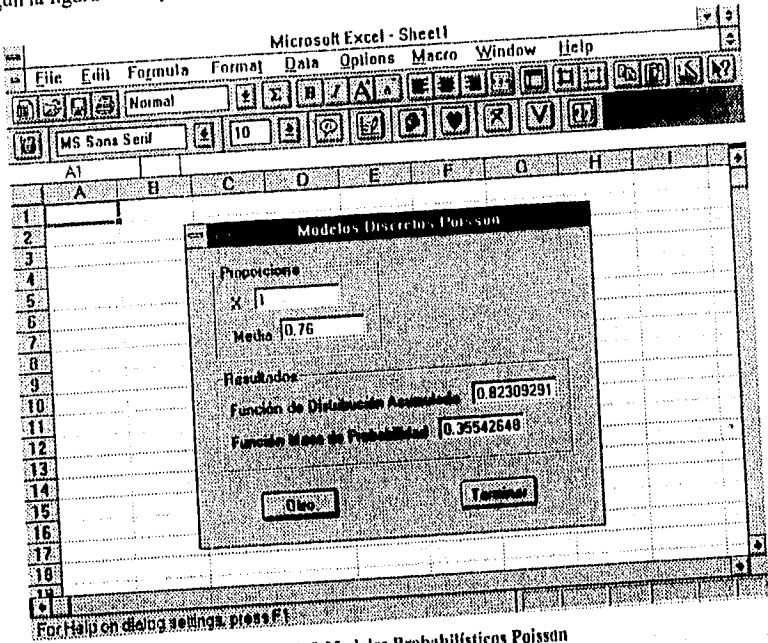


Figura 4.9 Modelos Probabilísticos Poisson

## Pruebas de Bondad de Ajuste

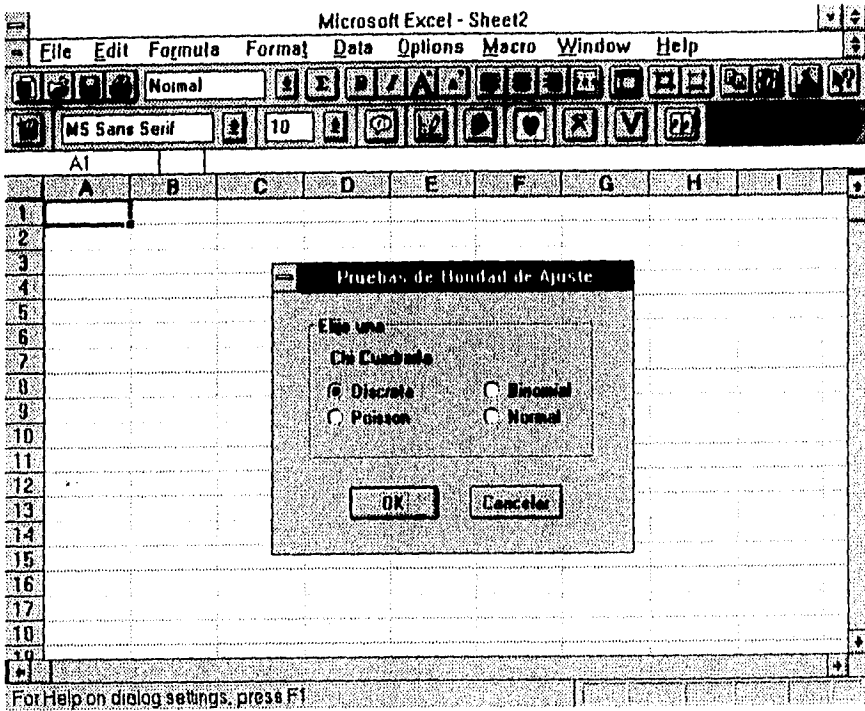


Figura 4.10 Menú Principal Bondad de Ajuste

### Chi Cuadrada Binomial

Un ingeniero que a menudo visita las Vegas a decidido simular 4 tornos automáticos mediante el lanzamiento de cuatro dados. El número de dados que resulten pares representará el número de tornos inactivos. El ingeniero planea lanzó cuatro dados un total de 50 veces. Los datos observados son los siguientes:

Número Dados Pares	0	1	2	3	4	Total
Frecuencia	4	13	16	13	4	50

¿Que tan bien los datos observados se asemejan a frecuencias binomiales?

Por lo que se plantea las siguientes hipótesis:

$H_0$  = La muestra tiene una distribución binomial

$H_1$  = La muestra no tiene una distribución binomial

La Figura 4.11 muestra la ventana diálogo.

El resultado de esta ejecución se incluye en la página 113.

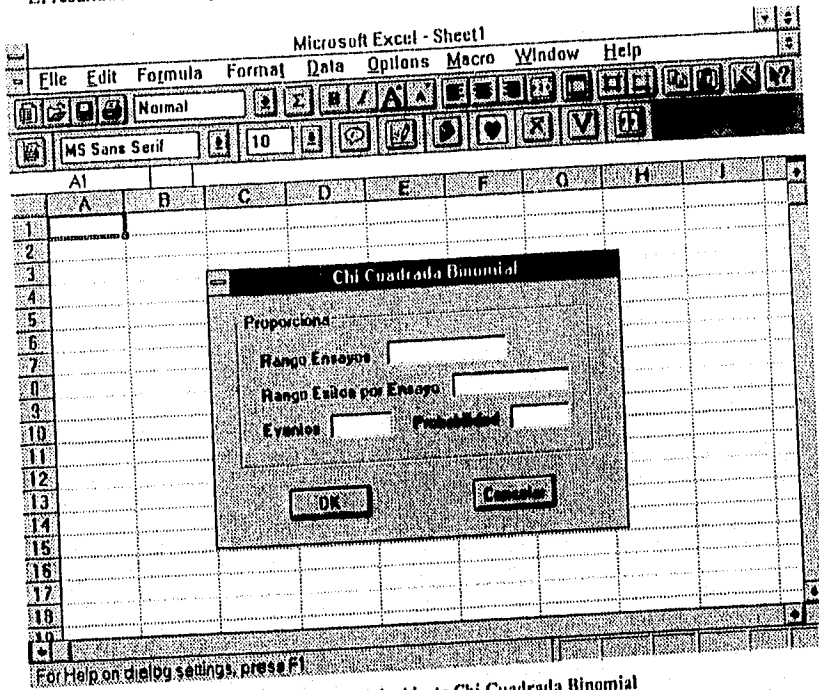


Figura 4.11 Bondad de Ajuste Chi Cuadrada Binomial



**Bondad de Ajuste**  
**Chi Cuadrada Binomial**

Ocurrencia	Ensayos	Frecuencias Observadas	Probabilidad Binomial	Frecuencias Esperadas	Chi Cuadrada
1	0	4	0.0625	3.125	0.245
2	1	13	0.25	12.5	0.02
3	2	16	0.375	18.75	0.4033
4	3	13	0.25	12.5	0.02
5	4	4	0.0625	3.125	0.245

k 5 Chi Cuadrada <= Chi Cuadrada Cero  
n 4 Per lo tanto se acepta la hipótesis  
Total de Frecuencias 50  
El valor de Chi Cuadrada es: 0.9333  
Grados de Libertad (k-1): 4  
Nivel de Significancia 0.05  
El valor de Chi Cuadrada Cero es: 9.4877

### Chi Cuadrada Discreta

El Director de Producción de Sealed Power Mexicana S.A., según reportes de productividad diaria, ha detectado los niveles de ausentismo de los lunes son el doble que los del resto de la semana. Por lo tanto ha solicitado a el Gerente de Planta que investigue que tan cierto es esto.

El Jefe de Personal de la Planta tiene registro de los últimos tres meses:

Día	Lunes	Martes	Miércoles	Jueves	Viernes
Días Perdidos	304	176	139	141	130

Si la razón de ausentismo es 2:1:1:1:1, la suma es 6. Por lo tanto, la función de probabilidad será:

$$f(a) = \begin{cases} 2/6 & i = 1 \\ 1/6 & i = 2,3,4,5 \end{cases}$$

Por lo que se plantea la siguiente hipótesis:

$H_0$  = La muestra tiene una distribución  $f(a)$

$H_1$  = La muestra no tiene una distribución  $f(a)$

El resultado de esta ejecución se incluye en la página 115.

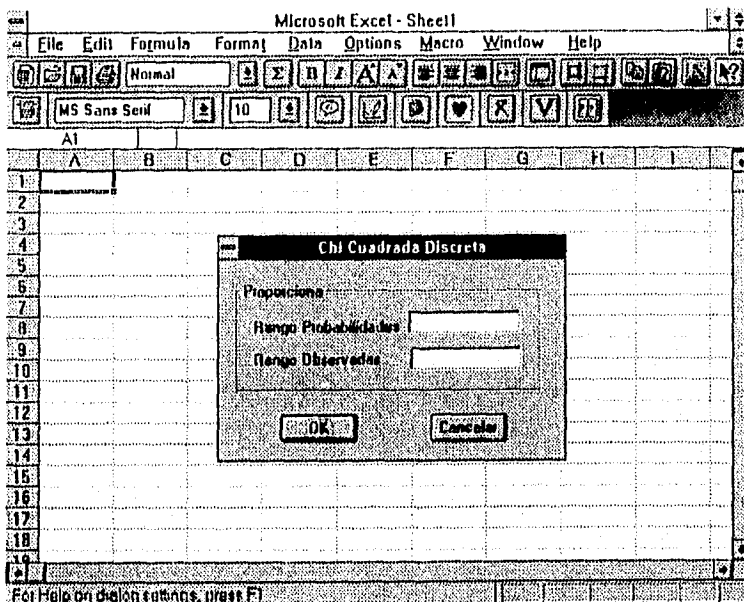


Figura 4.12 Chi Cuadrada Discreta

**Bondad de Ajuste**  
**Chi Cuadrada Discreta**

<u>Ocurrencia</u>	<u>Probabilidades</u>	<u>Observadas</u>	<u>Esperadas</u>	<u>Chi Cuadrada</u>
1	0.333333333	304	296.6667	0.1813
2	0.166666667	176	148.3333	5.1503
3	0.166666667	139	148.3333	0.5873
4	0.166666667	141	148.3333	0.3625
5	0.166666667	130	148.3333	2.2659
		-----	-----	-----
		890	889.9999	8.5573

k 5  
n 890  
El valor de Chi Cuadrada es: 8.5573  
Grados de Libertad (k-1): 4  
Nivel de Significancia 0.05  
El valor de Chi Cuadrada Cero es: 9.4877

Chi Cuadrada <= Chi Cuadrada Cero  
 Por lo tanto se acepta la hipótesis

### Chi Cuadrada Normal

Técnicos de la Comisión Nacional del Agua obtuvieron mediciones por cerca de tres meses, y ahora intentan ajustara una distribución Normal los datos que obtuvieron. Las 80 mediciones de flujo fueron registradas en un proceso monitoreado por computadora.

Los flujos representan valores medios y son los siguientes:

Flujos (L/s)	7.65	7.95	8.25	8.55	8.85	9.15
Frecuencia	5	21	35	15	3	1

Plantean las siguientes hipótesis:

$H_0$ : El flujo tiene una distribución normal.

$H_1$ : El flujo no tiene una distribución normal

La Figura 4.13 muestra la ventana diálogo.

El resultado de esta ejecución se incluye en la página 117.

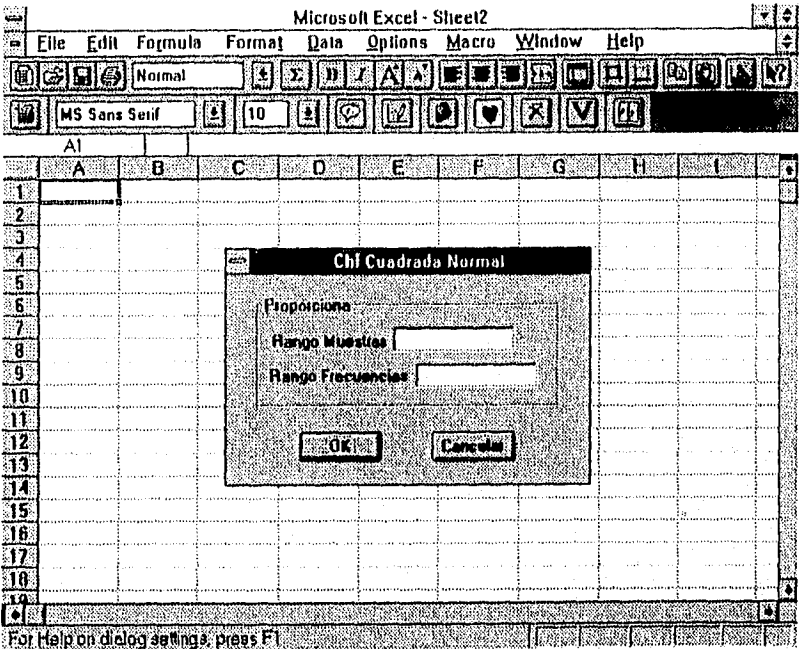


Figura 4.13 Bondad de Ajuste Normal

**Bondad de Ajuste**  
**Chi Cuadrada Normal**

Ocurrencia	Muestras	Frecuencias Observadas	Media	Desviación Estándar	Límite Superior x
1	7.65	5	8.2237	0.295	7.7975
2	7.95	21			8.0975
3	8.25	35			8.3975
4	8.55	15			8.6975
5	8.85	3			8.9975
6	9.15	1			9.2975

z	Area izquierda de z	Probabilidad	Frecuencia Esperada	Chi Cuadrada
-1.4447	0.0743	0.0743	5.94	0.1488
-0.4278	0.3344	0.2601	20.81	0.0017
0.5892	0.7221	0.3877	31.02	0.5107
1.6061	0.9459	0.2238	17.9	0.4698
2.6231	0.9956	0.0497	3.98	0.2413
3.64	0.9999	0.0043	0.34	1.2812

<u>n</u>	6
<u>Total de Frecuencias</u>	80
<u>El valor de Chi Cuadrada es:</u>	2.65
<u>Grados de Libertad (k-3)</u>	3
<u>Nivel de Significancia</u>	0.05
<u>El valor de Chi Cuadrada Cero es:</u>	7.8147

Chi Cuadrada <= Chi Cuadrada Cero  
Por lo tanto se acepta la hipótesis

### Chi Cuadrada Poisson

La planta de producción de Goodyear Oxo ha detectado problemas de contenido de azufre en la fabricación de llantas. El Jefe de Planta a solicitado a Gabriel Pérez, un ingeniero químico, que solucionara esto. Gabriel encontró que hasta hoy el límite del 4% de contenido de azufre era excedido 5 veces por la línea de producción A y 4 veces por la línea B.

En una muestra de 100 días, el número de días que fue violado límite del 4% en cada llanta es el siguiente:

Violaciones por día	0	1	2	3	4	5	6	Tot Frec
								115
Días	33	44	10	5	5	2	1	Tot Días
								100

Se utilizó la función de densidad de probabilidad Poisson y  $\lambda = \bar{X} = 1.15$ . Gabriel duda si Poisson fue una buena elección, por lo que se plantea la siguiente hipótesis:

$H_0$  = La muestra tiene una distribución Poisson

$H_1$  = La muestra no tiene una distribución Poisson

El resultado de esta ejecución se incluye en la página 119.

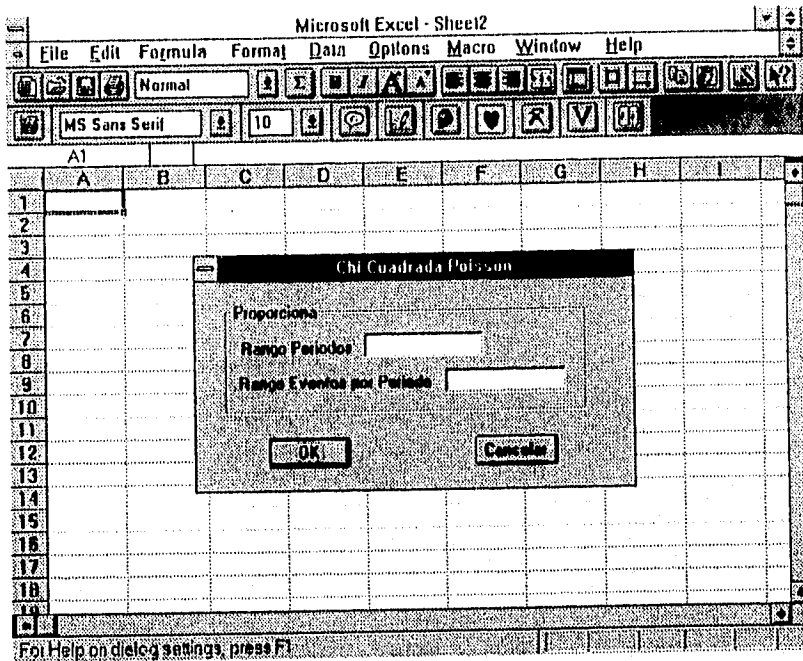


Figura 4.14 Bondad de Ajuste Poisson

**Bondad de Ajuste***Chi Cuadrada Poisson*

<u>Ocurrencia</u>	<u>Eventos Período</u>	<u>Periodos (Observados)</u>	<u>Probabilidad Poisson</u>	<u>Esperadas</u>	<u>Chi Cuadrada</u>
1	0	33	0.3166	31.66	0.0567
2	1	44	0.6808	36.41	1.5822
3	2	10	0.8901	20.94	5.7155
4	3	5	0.9704	8.03	1.1433
5	4	5	0.9935	2.31	3.1325
6	5	2	0.9988	0.53	4.0772
7	6	1	0.9998	0.1	8.1

<u>k</u>	7
<u>n</u>	100
<u>Lambda</u>	1.15
<u>El valor de Chi Cuadrada es:</u>	23.9074
<u>Grados de Libertad (k-1):</u>	5
<u>Nivel de Significancia</u>	0.05
<u>El valor de Chi Cuadrada Cero es:</u>	11.0705

**Chi Cuadrada > Chi Cuadrada Cero**  
**Por lo tanto se rechaza la hipótesis**

## Pruebas de Hipótesis Medias

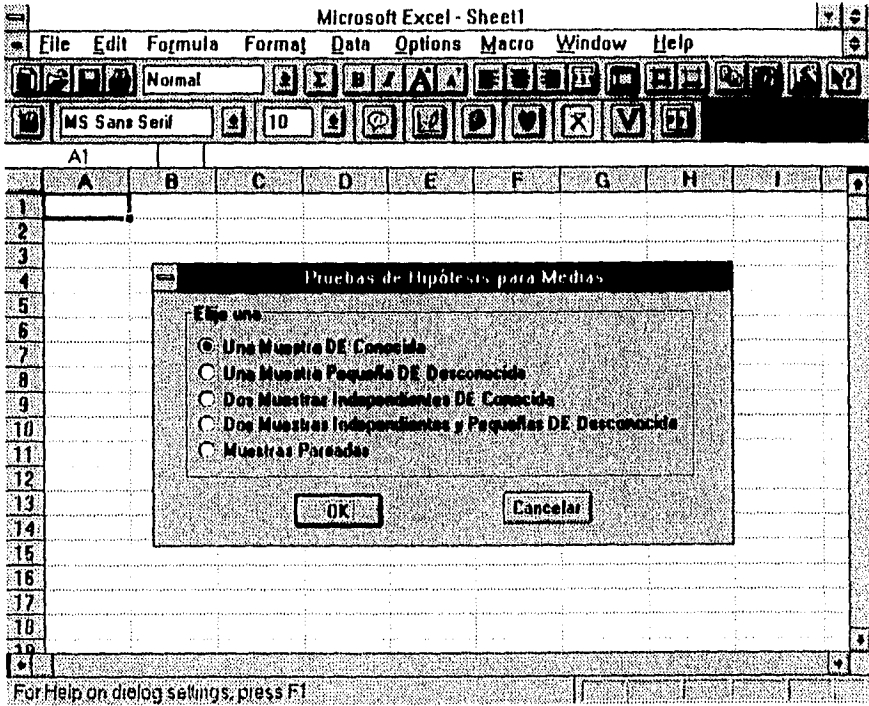


Figura 4.15 Menú Principal Prueba de Hipótesis Medias



### Una Muestra Desviación Estándar Conocida

La Planta de Manufactura de IBM de México en el Salto, Jalisco, está estudiando sus tiempos de proceso de la línea de producción de minicomputadoras AS/400.

Se tiene registrado que el proceso de ensamble tiene un tiempo estándar de 80 min. Después de algunos cambios en la línea, un aprendiz ensambla 10 productos en un tiempo promedio de 76.5 min. cada uno y con desviación estándar calculada de 6 min.

El departamento de aseguramiento de calidad planea si los promedios de tiempo son iguales, para lo que realizará una prueba de hipótesis de igualdad de medias. La hipótesis que se plantean es:

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$$

La muestra de los 10 productos es:

70	70	73	79	77
80	72	83	84	77

El resultado de esta ejecución se incluye en la página 122.

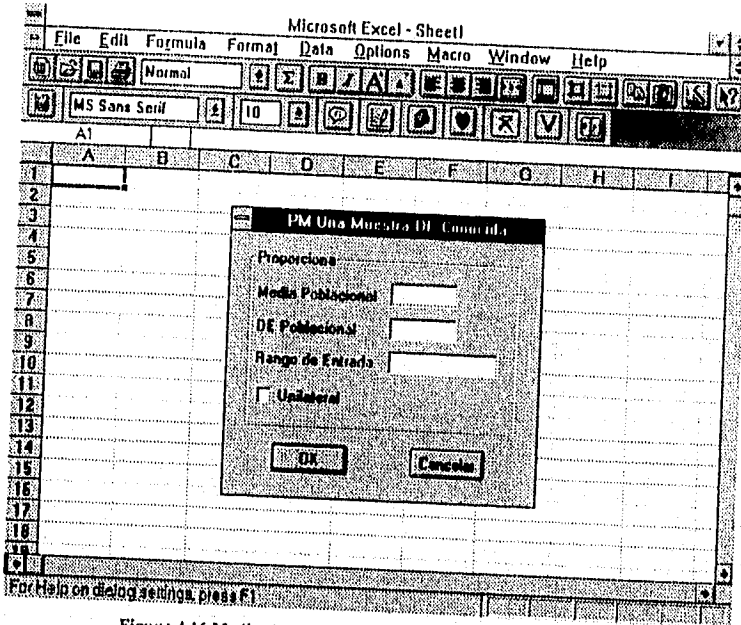


Figura 4.16 Medias Una Muestra Desviación Estándar Conocida

**Prueba de Medias**  
**Una Muestra Desviación Estándar Conocida**

Ocurrencia	Muestra	Unilateral	Media Poblacional	80
1	70	No	Media Muestral	76.5
2	80		Desviación Estándar	6
3	70			
4	72			
5	73			
6	83			
7	79			
8	84			
9	77			
10	77			

El valor de Z es: -1.8447  
El valor de Z Uno es: -1.96  
El valor de Z Dos es: 1.96

H Cero:  $\mu = \mu_0$   
H Uno:  $\mu < > \mu_0$   
El valor de Z está dentro del intervalo  $Z_{Uno} \leq Z \leq Z_{Dos}$   
Por lo tanto se acepta H Cero

### Una Muestra Pequeña Desviación Estándar Desconocida

Duracell S.A. de C.V. a través de su Departamento de Control de Calidad de su Planta Naulcalpan, ha reportado que el número de fallas observado de 5 baterías prototipo ha sido de 32, 41, 42, 49 y 53 hrs. La garantía de fábrica, establecida por la Dirección de Producción y fijada por competidores, es de 50 hrs. de funcionamiento.

¿Parecería que estas muestras pertenecen a una población con media de 50? Para verificar esto realizarán una prueba de hipótesis para lo que se plantea lo siguiente:

$$H_0: \mu = 50$$

$$H_1: \mu \neq 50$$

El resultado de esta ejecución se incluye en la página 124.

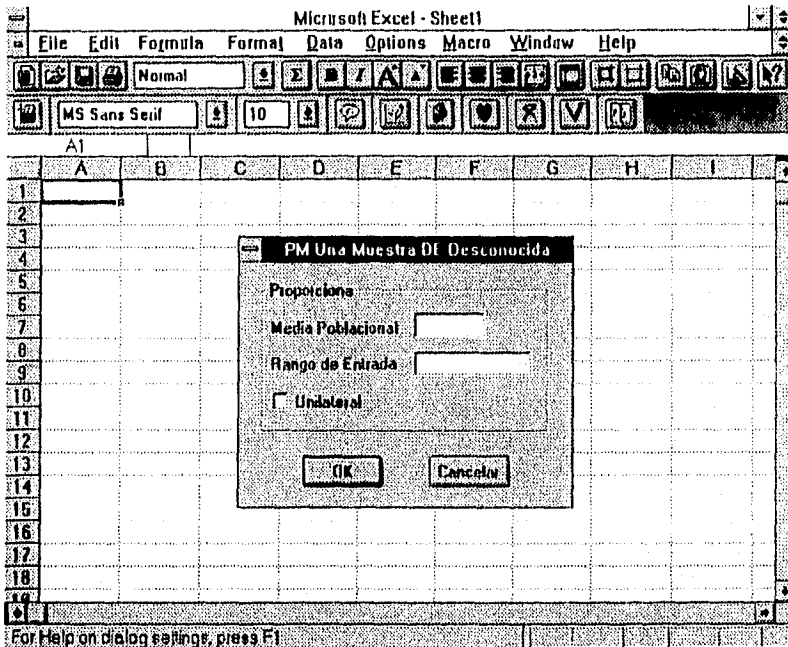


Figura 4.17 Medias Una Muestra Pequeña Desviación Estándar Desconocida

**Prueba de Medias**  
**Una Muestra Desviación Estándar Desconocida**

Ocurrencia	Muestra	Unifateral	Media Poblacional	50
1	32	No	Media Muestral	43.4
2	41		Desviación Estándar	8.0808
3	42			
4	49			
5	53			

El valor de t es: -1.8263      H Cero:  $\mu = \mu_{Cero}$   
El valor de t Uno es: -2.131846486      H Uno:  $\mu < > \mu_{Cero}$   
El valor de t Dos es: 2.131846486      El valor de t está dentro del intervalo  $t_{Uno} \leq t \leq t_{Dos}$   
Por lo tanto se acepta H Cero

### Dos Muestras Independientes Desviación Estándar Conocida

Dos ingenieros industriales que trabajan para la Gerencia de Calidad de Plásticos Mexicanos S.A. desean comparar los promedios de resistencia de plástico fabricado con procesos similares en dos plantas diferentes. Los procesos han estado bien establecidos y los valores  $\sigma$  son conocidos. Los datos son recopilados son:

Planta 1			Planta 2			
39	40	33	37	35	30	34
42	38	41	33	35	40	36
36	45	37	39	34	32	33
			31	36	38	37

¿Son iguales las medias? Para comprobar esto han decidido realizar una prueba de hipótesis de medias y han planteado la siguiente hipótesis

$$H_0: \mu_1 = 35 \text{ lb/in}^2$$

$$H_1: \mu_1 \neq 35 \text{ lb/in}^2$$

El resultado de esta ejecución se incluye en la página 126.

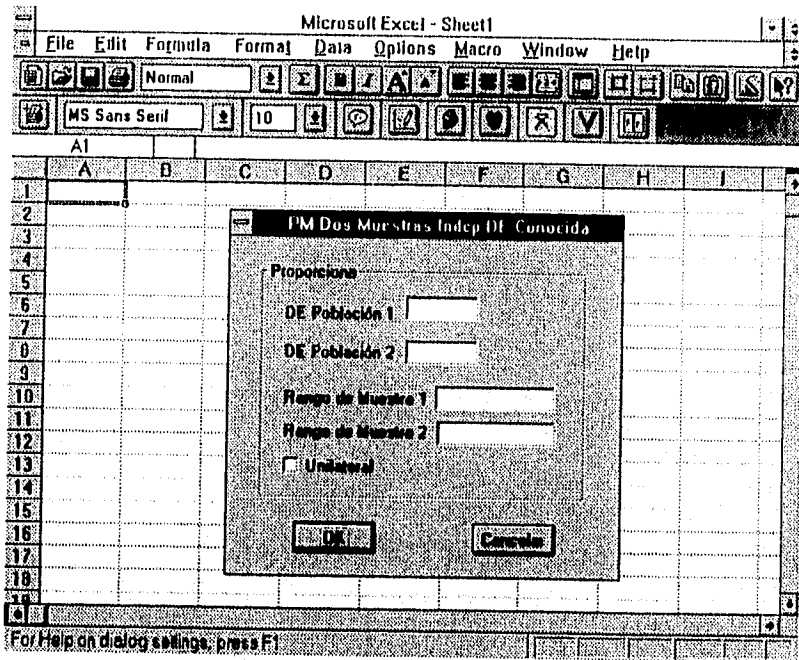


Figura 4.18 Medias Das Muestras Independientes Desviación Estándar Conocida

**Prueba de Medias**  
**Dos Muestras Independientes**  
**Desviación Estándar Conocida**

Ocurrencia 1	Muestra 1	Ocurrencia 2	Muestra 2	Unilateral	Media Muestral 1	
1	39	1	37	No	Desviación Estándar 1	3
2	42	2	33		Media Muestral 2	35
3	36	3	39		Desviación Estándar 2	5
4	40	4	31		D. E. de la diferencia	1.6008
5	38	5	35			
6	45	6	35			
7	33	7	34			
8	41	8	36			
9	37	9	30			
		10	40			
		11	32			
		12	38			
		13	34			
		14	36			
		15	33			
		16	37			

El valor de Z es: 2.4988 H Cero:  $\mu_1 = \mu_2$   
 El valor de Z Uno es: -1.96 H Uno:  $\mu_1 <> \mu_2$   
 El valor de Z Dos es: 1.96 El valor de Z no está dentro del intervalo  $Z_{Uno} \leq Z \leq Z_{Dos}$   
Por lo tanto se rechaza H Cero

### Dos Muestras Independientes y Pequeñas Desviación Estándar Desconocida

Cinemark de México con objeto de comparar la asistencia a sus salas de menos capacidad de asientos en México D.F. y Hermosillo, intenta comparar dos muestras de asistencia diaria por semana por lo que dos muestras de medias serán probadas por igualdad.

Las muestras recopiladas son:

Muestra 1 México  
53 55 58 50

Muestra 2 Hermosillo  
55 63 54 64

Si las medias son:

$$\bar{X}_1 = 54 \quad \bar{X}_2 = 59$$

La hipótesis que se plantea es:

$$H_0: 54 = 59$$

$$H_1: 54 \neq 59$$

El resultado de esta ejecución se incluye en la página 128.

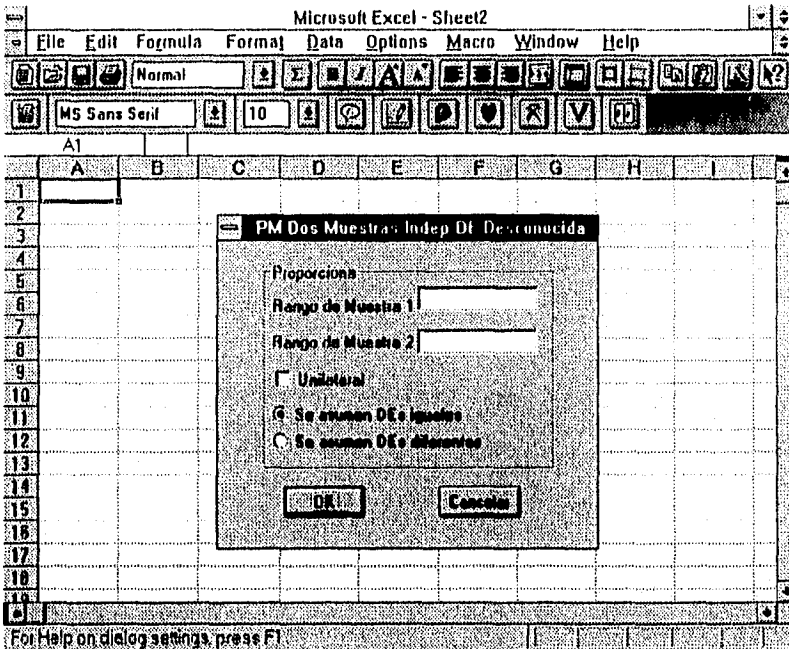


Figura 4.19 Dos Muestras Independientes y Pequeñas Desviación Estándar Desconocida

**Prueba de Medias**  
**Dos Muestras Independientes**  
**Desviación Estándar Desconocida**

Ocurrencia 1	Muestra 1	Ocurrencia 2	Muestra 2	Unilateral	Media Muestral 1	54
1	53	1	55	No	Desviación Estándar 1	3.3665
2	55	2	63		Media Muestral 2	59
3	58	3	54		Desviación Estándar 2	5.2281
4	50	4	64			

El valor de t es: -1.6082 H Cero:  $\mu_{Uno} = \mu_{Dos}$   
 El valor de t Uno es: -2.446914 H Uno:  $\mu_{Uno} < > \mu_{Dos}$   
 El valor de t Dos es: 2.446914 El valor de t está dentro del intervalo  $t_{Uno} \leq t \leq t_{Dos}$   
 Nivel de Significancia: 0.5 Por lo tanto se acepta H Cero  
 Grados de Libertad: 6



## Dos Muestras Pareadas

CLEVITE de México S.A., proveedor Q1 de Ford Motor Company, ha desarrollado un cojinete experimental que debe ser tratado con procesos a base de calor. Para poder determinar cuando dos métodos son equivalentes, un ingeniero metalúrgico eligió 12 muestras y sometió cada una a temperaturas de 60 y 79 °C. La dureza de las piezas se obtuvieron los valores de las siguientes tablas.

¿Sería correcto afirmar que la temperatura no tuvo que ver con la dureza?

Plantea las siguientes hipótesis:

$$H_0: \delta = 0 \quad H_1: \delta \neq 0$$

El resultado de esta ejecución se incluye en la página 130.

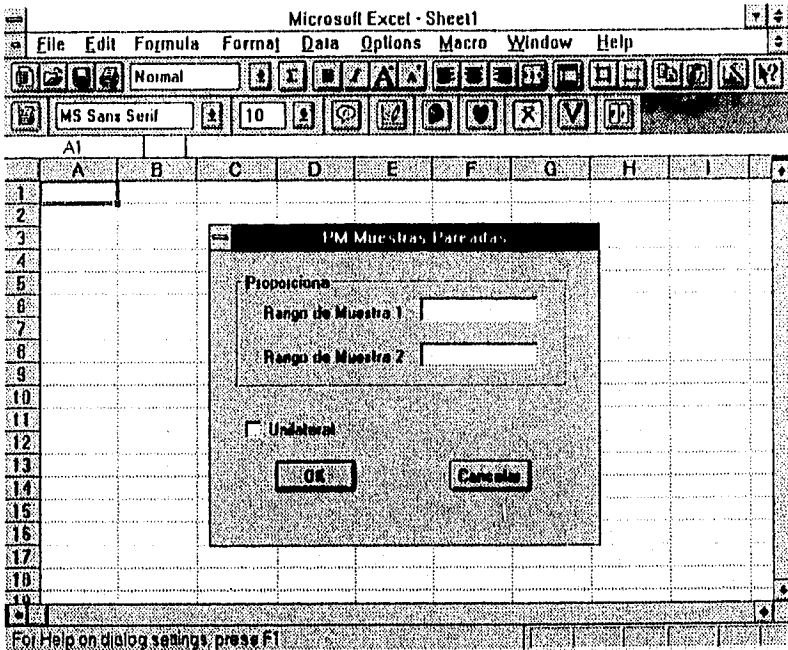


Figura 4.20 Dos Muestras Pareadas

**Prueba de Medias  
Dos Muestras Pareadas**

<u>Ocurrencia</u>	<u>Muestra 1</u>	<u>Muestra 2</u>	<u>Unilateral</u>	<u>Diferencia di</u>	<u>Diferencia di cuadrada</u>
1	81	76	No	5	25
2	79	79		0	0
3	82	80		2	4
4	79	75		4	16
5	83	77		6	36
6	80	74		6	36
7	83	76		7	49
8	81	80		1	1
9	80	82		-2	4
10	77	79		-2	4
11	74	75		-1	1
12	81	81		0	0
				26	176

Media d: 2.1667  
Desviación Estándar Sd: 3.2983  
Nivel de Significancia: 0.5  
Grados de Libertad: 11

El valor de t es: 2.2756  
El valor de t Uno es: -2.201  
El valor de t Dos es: 2.201

H Cero: delta = 0  
H Uno: delta <> 0  
El valor de t no está dentro del intervalo t Uno <= t <= t Dos  
Por lo tanto se rechaza H Cero

## V Pruebas de Hipótesis Variancias

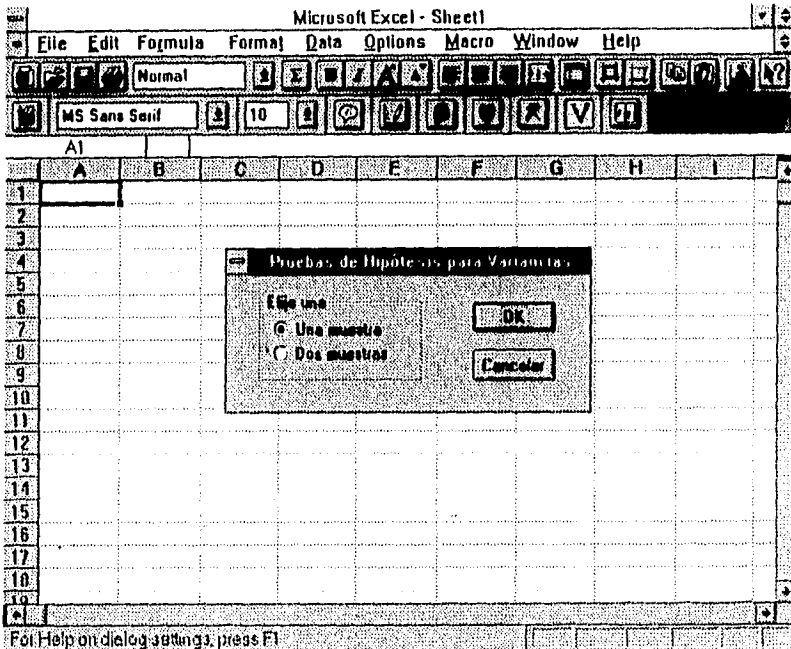


Figura 4.21 Menú Principal Prueba de Hipótesis Variancias

### Una Muestra

La Gerencia de Investigación y Desarrollo de IUSA S.A. ha detectado que ciertos prototipos de tubos de acero inoxidable de 20 cm de diámetro, tienen un flujo que presentan una variancia de 3  $\%$ . Un nuevo diseño se probó 30 veces, si la variancia muestral fue 2.95  $\%$ , se ha solicitado al Departamento de Diseño que investigue si las variancias son iguales.

La muestra de donde se obtuvo la información anterior es la siguiente:

Muestra				
32	32	32	35	30
31	31	29	29	29
29	29	30	30	30
29	28	28	29	31
32	31	33	30	29
28	33	32	31	32

Las hipótesis que se plantean son:

$$H_0: \sigma^2 = 3 \quad H_1: \sigma^2 > 3$$

El resultado de esta ejecución se incluye en la página 133.

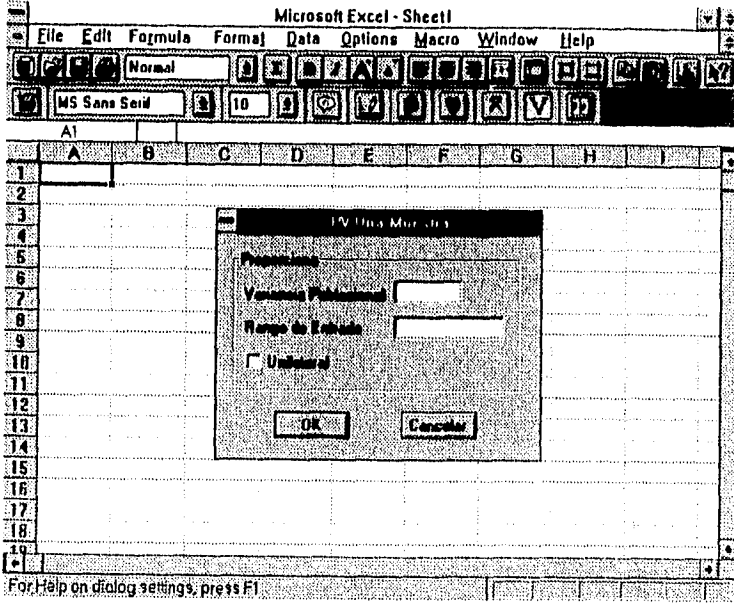


Figura 4.21 Variancias Una Muestra

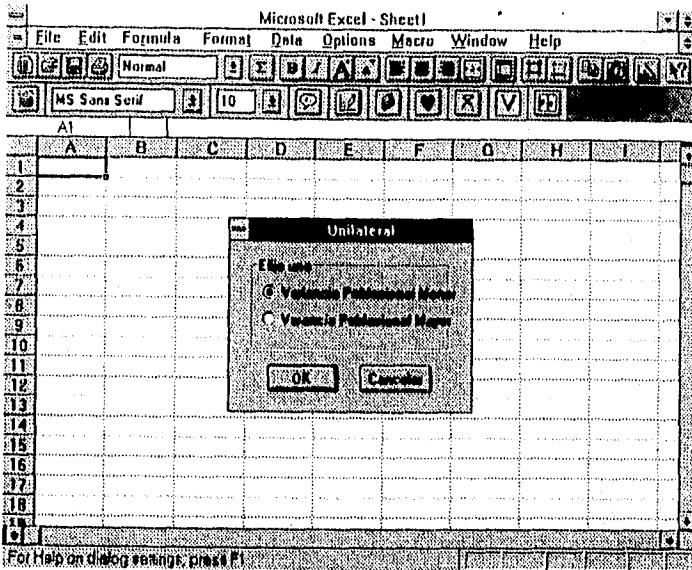


Figura 4.22 Variancias Un Muestra Unilateral

**Prueba de Variancias**  
**Una Muestra**

Ocurrencia	Muestra	Variancia Poblacional	Unilateral	Variancia Muestral	2.947126437
1	32	3	Si	<u>Desviación Estándar</u>	1.716719673
2	31			<u>Nivel de Significancia</u>	0.5
3	29		<u>Variancia Poblacional</u>	<u>Grados de Libertad</u>	29
4	29		Menor		
5	32				
6	28				
7	32				
8	31				
9	29	<u>El valor de Chi2 es</u>	28.4889	<u>H_Cero: Sigma2 = Sigma2_Cero</u>	
10	28	<u>El valor de Chi2_Dos es</u>	42.5569	<u>H_Uno: Sigma2 &lt; Sigma2_Cero</u>	
11	31			<u>Puesto que Chi2 &lt;= Chi2_Dos</u>	
12	33			<u>Por lo tanto se acepta H_Cero</u>	
13	32				
14	29				
15	30				
16	28				
17	33				
18	32				
19	35				
20	29				
21	30				
22	29				
23	30				
24	31				
25	30				
26	29				
27	30				
28	31				
29	29				
30	32				

## Dos Muestras

Los alumnos de la Facultad de Química prueban dos nuevos métodos de medición de temperatura de reacción química. Las muestras de tamaño 15 y 20 fueron tomadas y el tiempo para determinar la temperatura correcta es registrado. Las variancias son  $0.15 s^2$  para 15 elementos y  $0.20 s^2$  para 20 elementos. Y las muestras son las siguientes:

Muestra 1			
32	29	32	31
31	28	31	29
30	32	32	32
29	29	29	31
31	28	30	30

Muestra 2		
29	30	29
30	29	30
32	31	32
32	32	29
29	30	31

Las hipótesis que se plantean son:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

El resultado de esta ejecución se incluye en la página 135.

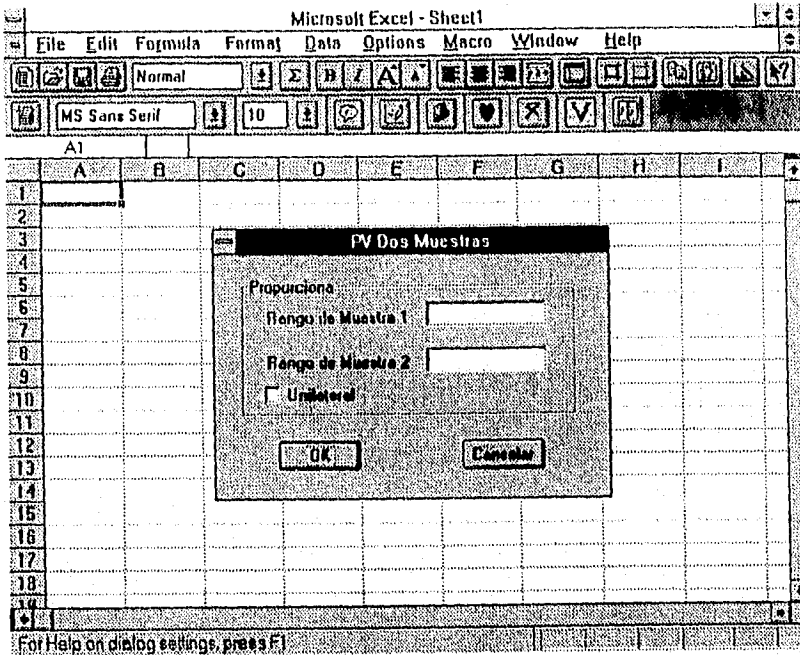


Figura 4.23 Variancias Dos Muestras

**Prueba de Variancias****Das Muestras**

Ocurrencia 1	Muestra 1	Ocurrencia 2	Muestra 2	Unilateral	Variancia Muestral 1	0.2
1	32	1	29	Si	Variancia Muestral 2	0.15
2	31	2	30		Grados de Libertad 1	19
3	30	3	32	Variancia 1	Grados de Libertad 2	14
4	29	4	32	Mayor	Nivel de Significancia	0.5
5	31	5	29			
6	29	6	30			
7	28	7	29			
8	32	8	31			
9	29	9	32			
10	28	10	30			
11	32	11	29			
12	31	12	30			
13	32	13	32			
14	29	14	29			
15	30	15	31			
16	31					
17	29					
18	32					
19	31					
20	30					

El valor de F es:

1.3333

El valor de F Uno es:

1

El valor de F Dos es:

1.9701

H Cero:  $\sigma_2 \text{ Uno} = \sigma_2 \text{ Dos}$ H Uno:  $\sigma_2 \text{ Uno} > \sigma_2 \text{ Dos}$ El valor de F está dentro del intervalo  $1 \leq F \leq F \text{ Dos}$ Por lo tanto se acepta H Cero

## Pruebas de Hipótesis Proporciones

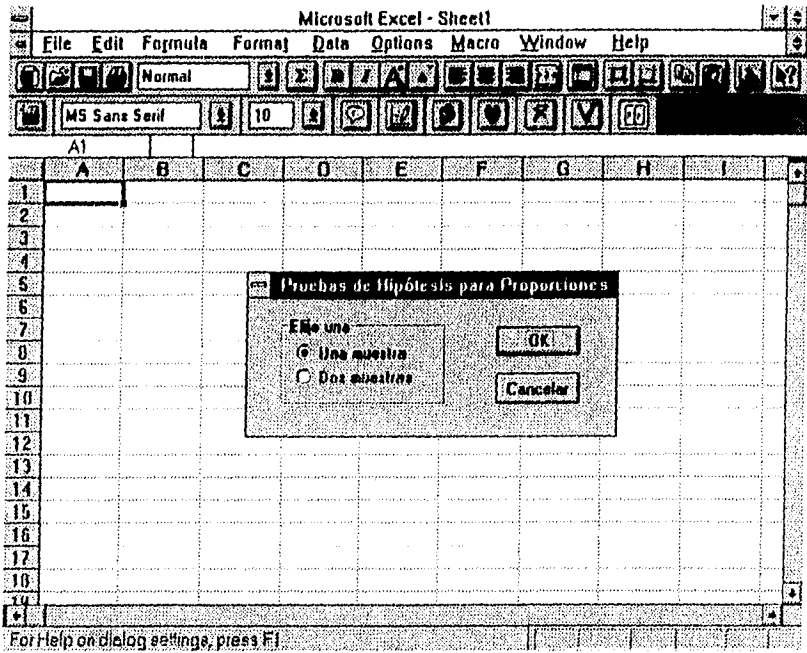


Figura 4.24 Menú Principal Prueba de Hipótesis Proporciones

### Una Muestra

Históricamente la fracción de productos defectuosos fabricados por la compañía Industrial del Norte S.A. ha sido de 0.02. A causa de la creciente demanda de sus productos se ha decidido subcontratar parte de la fabricación, por lo que fueron enviados 5 lotes a un contratista. De estos productos se tomó una muestra de 500 elementos y fue encontrado un promedio de 0.01 defectos. Estimulado por la posible obtención de un contrato más grande el contratista planeaba disminuir aun más esta proporción de defectuosos.

Antes de entregar un pedido más grande la Dirección de Industrial del Norte S.A. se pregunta si en realidad la calidad se ha mejorado.

Para contestar lo anterior el Departamento de Compras ha decidido realizar una prueba de hipótesis para proporciones con las siguientes hipótesis.

$$H_0: \pi = 0.02$$

$$H_1: \pi < 0.02$$

El resultado de esta ejecución se incluye en la página 138.



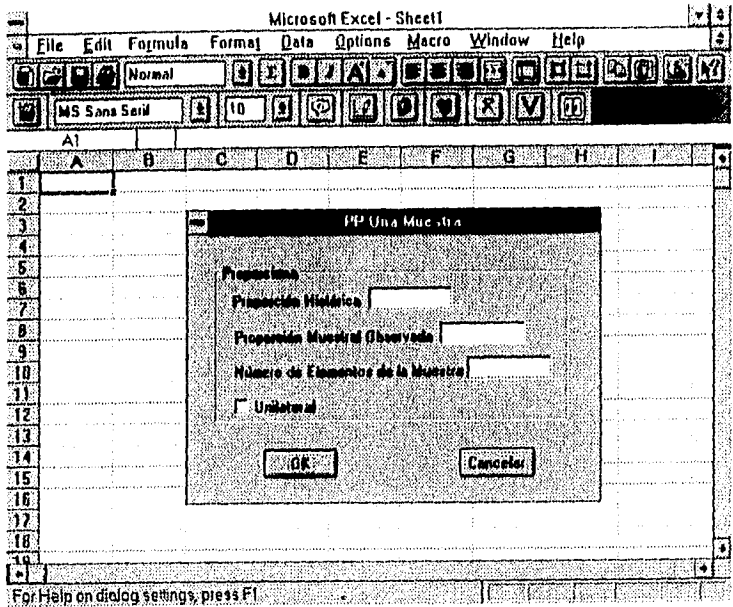


Figura 4.25 Proporciones Una Muestra

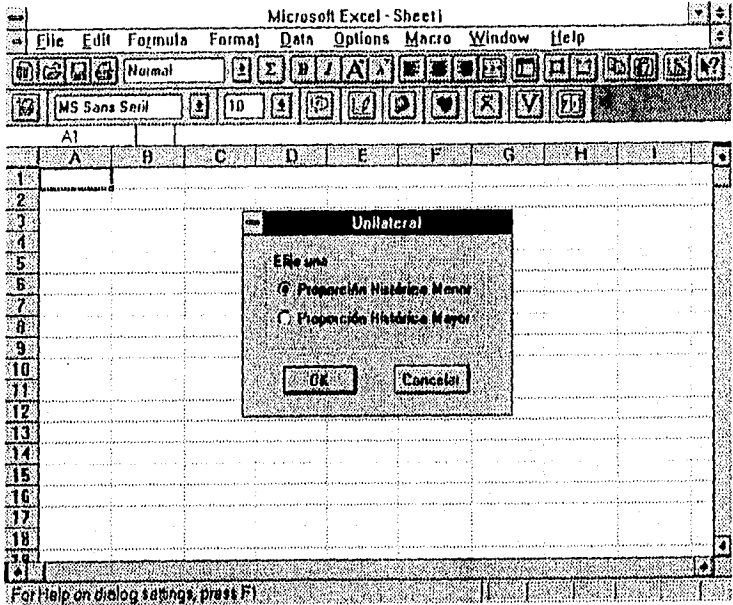


Figura 4.26 Proporciones Una Muestra Unilateral

**Prueba de Proporciones**

Una Muestra

<u>Proporción Histórica</u>	<u>Proporción Muestral</u>	<u>Unilateral</u>	<u>Desviación Estándar</u>	0.0063
0.02	0.01	Si	<u>Nivel de Significancia</u>	0.05
	<u>Elementos de la Muestra</u>	<u>Proporción Histórica</u>		
	500	Mayor		
<u>El valor de Z es</u>	-1.5972	<u>H Cero: <math>P_i = P_i \text{ Cero}</math></u>		
<u>El valor de Z uno es</u>	-1.645	<u>H Uno: <math>P_i &lt; P_i \text{ Cero}</math></u>		
		<u>Puesto que <math>Z \text{ Uno} \leq Z</math></u>		
		<u>Por lo tanto se acepta H Cero</u>		

## Dos Muestras

La División de Mejora Continua de Condumex S.A. busca reducir la tasa de defectos, para lo cual realiza el siguiente estudio.

Ocho metros de alambre de cobre han sido producidos en dos plantas diferentes con una especificación máxima de resistencia 1 ohm ( $\Omega$ ). Históricamente las cantidades de defectos en cada planta han sido estadísticamente iguales con un 10% en cada planta A y 11% en la planta B. Recientemente se obtuvieron muestras de 150 y 100 alambres de la planta A y B, respectivamente cada una registró 12 alambres que no cumplieron con esta especificación.

El Jefe de la División se cuestiona si las tasas de defectuosos son estadísticamente iguales. Se ha propuesto realizar la siguiente prueba de hipótesis:

$$H_0: \pi_1 = \pi_2 \quad H_1: \pi_1 \neq \pi_2$$

El resultado de esta ejecución se incluye en la página 140.

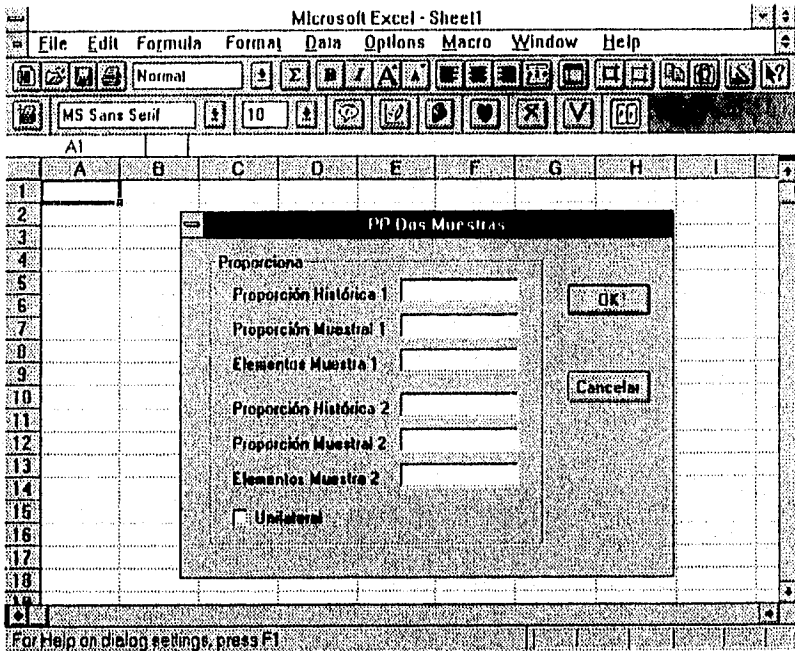


Figura 4.27 Proporciones Dos Muestras

**Prueba de Proporciones**  
 Dos Muestras

<u>Proporción Histórica 1</u>	<u>Proporción Muestral 1</u>	<u>Proporción Histórica 2</u>	<u>Proporción Muestral 2</u>	<u>Unilateral</u>	<u>Desviación Estándar (P1-P2)</u>	<u>0.0397</u>
0.1	0.08	0.11	0.12	No	<u>Nivel de Significancia</u>	0.05
	<u>Elementos Muestra 1</u>		<u>Elementos Muestra 2</u>			
	150		100			

<u>El valor de Z es:</u>	-1.0066	<u>H Cero: P1 Uno = P1 Dos</u>
<u>El valor de Z Uno es:</u>	-1.96	<u>H Uno: P1 Uno &lt;&gt; P1 Dos</u>
<u>El valor de Z Dos es:</u>	1.96	<u>El valor de Z está dentro del intervalo Z Uno &lt;= Z &lt;= Z Dos</u>
		<u>Por lo tanto se acepta H Cero</u>

## *Conclusiones y Recomendaciones*

A medida que pasa el tiempo y se desarrolla la tecnología de las computadoras, es posible afirmar que la creatividad y el conocimiento humano se verán cada vez más estimulados por esta disciplina.

Casi no existe ninguna actividad humana que no haya sido abordada por las computadoras. Un niño puede aprender francés en su habitación, haciendo uso de un programa de multimedia. En vez de construir un modelo físico, un ingeniero industrial puede ahorrar miles de pesos y cientos de horas, probando su prototipo en una computadora personal. En una misma noche un gerente de supermercado puede procesar inventarios, cambios de precios, ventas e inclusive encuestas de satisfacción de clientes.

Estos son sólo algunos ejemplos de lo que el hombre, en conjunción con la computadora, ha logrado.

Las aplicaciones de computadora que se han desarrollado, han revolucionado la forma de trabajar, de investigar, de estudiar y hasta de divertirse. Ciencias que siempre fueron dominadas sólo por expertos pueden ahora estar a disposición de quien necesite el apoyo de alguna especialidad del saber. Las computadoras pueden facilitar la ejecución de cálculos o de complicadas tareas, y de una manera sencilla proporcionar resultados e inclusive conclusiones; sólo serán necesarios ciertos conocimientos básicos y algo de interpretación de la información obtenida. Este podría ser el caso de la inferencia estadística, útil en los negocios, la ingeniería, la investigación, la administración, la medicina, la producción industrial y muchos otros aspectos del conocimiento.

Una de las finalidades de las computadoras es la optimización de recursos, este sistema tiene este como uno de sus principales objetivos, siendo el tiempo el recurso escaso. Otro objetivo de este sistema es aplicar y desarrollar las habilidades que a lo largo del desempeño académico y profesional se han adquirido.

El análisis, desarrollo e implantación de sistemas para computadora ha evolucionado vertiginosamente al paso del tiempo. Lo que también debe de evolucionar son las alternativas de desarrollo. No se debe de limitar a las aplicaciones tradicionales ni a las herramientas tradicionales, se deben de buscar nuevos ambientes, lenguajes e interfaces que faciliten el desarrollo de aplicaciones, que a su vez faciliten a otros el desempeño de sus labores. De igual importancia es el tema a desarrollar, existen sofisticados programas de computadora que realizan funciones complejas; pero, generalmente, estos consumen muchos recursos y su aplicación está orientada a especialistas. Si se desarrollan programas que hagan accesible el conocimiento a cada vez más personas, sea cual fuere su profesión, podemos decir que hemos logrado hacer una buena aportación como conocedores y practicantes de la computación.

Los Sistemas de Información se han convertido en un tópico de estrategia, han pasado de ser una fábricas de papel a elementos indispensables en el control y el desempeño de las organizaciones. El estudio y aprendizaje de las disciplinas relacionadas con la información traerá como consecuencia una mejor toma de decisiones y por consecuencia, un mejor y menos incierto futuro.

La experiencia y la historia han demostrado que para que una nación se enfile hacia el desarrollo sostenido, son indispensables la educación, el trabajo y la tecnología. Por lo que quienes hemos tenido la libertad y el privilegio de poscerlos, tenemos también la obligación de aplicarlos y difundirlos.

## *Bibliografía*

Blank, Leland  
*Statistical Procedures for Engineering, Management and Science*  
Mc Graw Hill, 1980

Borrás, H., Iriarte R. & Fontana, B.  
*Apuntes de Probabilidad y Estadística*  
Facultad de Ingeniería UNAM, 1985

Kinata, C. & Kyd, C.W.  
*Complete Guide to Microsoft Excel Macros*  
Microsoft Press, 1993

Kreyszig, Erwin  
*Introducción a la Estadística Matemática*  
LIMUSA, 1985

Lucas, Henry C.  
*The Analysis, Design and Implementation of Information Systems*  
Mc Graw Hill, 1985

*Microsoft Excel Function Reference*  
Microsoft Corporation, 1992

*Microsoft Excel User's Guide 1 & 2*  
Microsoft Corporation, 1992

Pressman, Roger S.  
*Software Engineering, 2nd Ed.*  
Mc Graw Hill, 1988

Saad, Miguel A.  
*Redacción*  
CECSA, 1982

Wexler S. & Sharer, J.  
*Microsoft Excel Macros. Step by Step*  
Microsoft Press, 1992