



39

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES "ACATLÁN"

SELECCIÓN Y USO DE SOFTWARE ESTADÍSTICO PARA EL MANEJO DE DATOS EXPERIMENTALES (EL CASO DE UNA ESTACIÓN METEOROLÓGICA DEL BACHILLERATO)

INFORME DE PRACTICA PROFESIONAL AL SERVICIO DE LA
COMUNIDAD

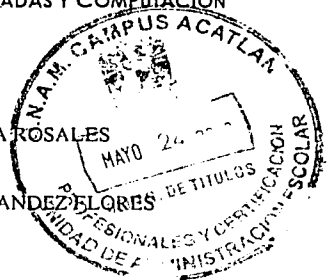
QUE PARA OBTENER EL TITULO DE.

LICENCIADA EN MATEMÁTICAS APLICADAS Y COMPUTACIÓN

PRESENTA

ROSA LAURA VERGARA ROSALES

ASESOR: DR. RAFAEL FERNANDEZ FLORES



MAYO 2002

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

AGRADEZCO :

Al amor de mi vida, DIOS, por todo lo que tengo inmerecidamente, por permitirme conocerlo y vivirlo, y por ...

Al hombre más importante de mi vida, mi papá, Jesús Vergara González; aunque tus ojos no pueden ver esto, siempre estarás presente con tus enseñanzas predicadas con ejemplos de amor, valentía, Intelligencia y honestidad.

A la mujer más importante de mi vida, mi mamá, María Estela Rosales Pérez; por entregar día con día, de manera ilimitada, valiente e incesante atenciones y apoyo, en una palabra, tu vida y amor, a cada uno de tus hijos. Gracias por aguantarme y animarme en los tiempos difíciles y fáciles también.

A mis queridos hermanos, David, Javier, Alberto, Eduardo, Teresa y Guadalupe, porque su vida ha sido un ejemplo para mí. Sin su ayuda no hubiera podido seguir; especialmente a Alberto a quien no me alcanza el papel para darle gracias por tanto que nos ayuda. Y de igual manera, les doy las gracias a Lourdes Palacios, Victoria Moreno, Martha Lamadrid y Gustavo Resendiz por todas sus atenciones y tolerancia. La agradable compañía de mi Tía Tere en las buenas y malas.

La alegría desparramada y escandalosa que he recibido de mis sobrinos, Alicia, David, Dany, Gaby, Lallito, Yoali, Luisito, Viridiana, Montserrat y Everlin; sin olvidar a mis nietas Samanta y Andrea.

A mi director de tesis, Dr. Rafael Fernández Flores, por sus enseñanzas, paciencia, guía y comprensión durante la realización de este trabajo así como por todas las oportunidades que me ha dado; al Dr. Ernesto Jauregul Ostos por los consejos y atenciones que me prestó; a la Lic. Ligia Kamms por su amabilidad y tanto apoyo. Sin ustedes hubiera sido imposible realizar este trabajo.

A mis amigos, los de antes y los de siempre, muy especialmente a mi amiga Guadalupe de la Rosa, por comprender y respetar; a Guadalupe Díaz por que ni la distancia ni el tiempo nos hacen cosquillas, a mi inseparable confidente y compañero Arturo Reyes, a Eduardo Flores, Soledad Ramirez, Adela y Rocío Castro, Sara Romero, Viola Gaspar, Mario Granados, Gloria García, Maricela Gomez, Ernesto y Adrián Muñiz, Jorge L. Esquivel y Carmen Resendiz, por compartir conmigo muy buenos momentos.

La lección de fe y lucha que he recibido de una mujer extraordinaria, María Luisa Bernal, también doy gracias por su cariño a Lucy Perez y Edith, y aunque sea tarde a mi amigo inolvidable el Sr. Leonardo. Gracias Araceli por el ejemplo y por que sin ti no habría llegado a ellos.

Por su ejemplo de entrega y compromiso a todos mis profesores, especialmente a los catedráticos Clavel, Madariaga, Reyes, Zamudio, Romero y Palencia de Acatlán.

A las profesoras designadas como sinodales: Act. Luz María Lavín, Ing. Nora del Consuelo Goris, Ing. Sandra Ortiz y Lic. Guadalupe del Carmen Rodríguez por la atención dedicada a este trabajo.

Las facilidades para la realización e Impresión de este Informe a la Dirección General de Servicios de Cómputo Académico, el Instituto de Matemáticas y el Colegio de Ciencias y Humanidades Plantel Sur.

TABLA DE CONTENIDO

INTRODUCCIÓN	1
1. DESCRIPCIÓN DEL PROGRAMA DE SERVICIO SOCIAL	7
1.1 Antecedentes Sobre Análisis Estadísticos Atmosféricos en el Valle de México	11
1.2 Objetivos y Meta	15
2. SOFTWARE ESTADÍSTICO Y CRITERIOS DE SELECCIÓN	17
2.1 Software y su Clasificación	19
2.2 La Estadística y la Necesidad de su Utilización en Diversas Áreas de Conocimiento	21
2.3 Breve Historia del Software Estadístico	22
2.3.1 Los Programas Estadísticos Pioneros	22
2.3.2 La Generación Para Pc	24
2.3.3 Software Estadístico Hoy	25
2.4 Características Determinantes en la Elección de un Programa Estadístico	28
2.4.1 Contenido Estadístico y Gráfico de los Programas de Propósito General	28
2.4.2. Interface	30
2.4.2.1. Diálogo	30
2.4.2.2. Diseño de Pantalla	31
2.4.2.3. Retroalimentación y Ayuda	32
2.4.2.4. Control De Errores	32
2.4.3. Plataforma	33
2.4.4. Características Generales	34
2.4.5. Costo	34
3. DESCRIPCIÓN DE LAS ESTACIONES METEOROLÓGICAS Y LOS DATOS EXPERIMENTALES	39
3.1 Antecedentes	41
3.2 Algunos Conceptos Básicos de Meteorología	43
3.3 Documentación de la Estación	46
3.4 Descripción de los Instrumentos Meteorológicos	47
3.5 Descripción del Software de la Estación - Weatherlink	50
3.6 Actual Manejo Estadístico de los Datos	51

4. ANÁLISIS ESTADÍSTICO	53
4.1 Los Datos	55
4.1.1. Exportación e Importación de Datos	56
4.1.2. Validación	67
4.1.2.1. Pruebas de Identificación e Integridad	67
4.1.2.2. Pruebas de Consecuencia Interna	68
4.1.2.3. Pruebas de Tolerancia	69
4.1.2.4. Pruebas de Consecuencia Temporal	70
4.2 Estadística Descriptiva	84
4.2.1. Tablas de Frecuencia y su Representación Gráfica	87
4.2.2. Medidas Numéricas Descriptivas	98
4.2.3. Análisis Bivariado	102
4.2.3.1. Tablas de contingencia	102
4.2.3.2. Gráficas de dispersión	105
4.3 Estadística Inferencial	109
4.3.1. Distribuciones de probabilidad empíricas y teóricas (Variables Continuas)	110
4.3.2. Distribución Normal	112
4.3.3. Teorema del límite central	117
4.3.4. Distribución t de Student	118
4.3.5. Estimación del parámetro μ e Intervalos de confianza	119
4.3.6. Pruebas de hipótesis estadísticas	122
4.3.7. Pruebas no paramétricas	134
 RESULTADOS Y CONCLUSIONES	 139
 REFERENCIAS	 145
 ANEXOS	 155

INTRODUCCIÓN

INTRODUCCIÓN



Aman zan nitliconextlahli

A menudo se escucha la palabra "estadística" relacionada a cifras de nacimientos, impuestos, ingresos, deudas, salud, votaciones, deportes, etc., de lo cual se puede deducir que cada vez hay un mayor número de disciplinas que requieren de conocimientos sobre la ciencia llamada Estadística¹. Es por esto, que en los programas de estudio de diversas licenciaturas e incluso en el bachillerato es requisito un curso básico de dicha materia.

La estadística no es "elitista" pues no está al servicio de alguna clase social, ciencia o campo de aplicación en particular, por el contrario, pone a la disposición, tanto de las ciencias físicas o naturales como de las sociales, un valioso conjunto de métodos que auxilian en el análisis de los datos obtenidos en sus respectivas investigaciones, tanto en lo referente a la estadística descriptiva (para delinear las características de la muestra o población), como a la inferencial (para probar hipótesis o estimar las características reales de la población de estudio).

Recordemos que las primeras generaciones de computadoras eran creadas con el propósito de ser instrumentos de cálculo y gracias a esto las matemáticas y todas sus ramas, como la estadística, se vieron beneficiadas ya que antes de que las computadoras existieran, los estadísticos tenían grandes problemas para el desarrollo óptimo de sus investigaciones; durante el siglo XIX se desarrollaron teorías que no pudieron ser calculadas adecuadamente debido a las herramientas con las que se contaba en esos momentos, incluso hasta mediados del siglo pasado los cálculos estadísticos complejos que se requerían se realizaban gracias a muchas personas que apoyándose en las calculadoras primitivas trabajaban con una gran cantidad de datos.

Por ello, cuando las computadoras estuvieron a la disposición de los estadísticos, estos se apoderaron rápidamente de dicha herramienta. A la par de los desarrollos del hardware, el software estadístico - y cualquier otro -, ha tenido que adecuarse a diversas transiciones tales como el paso de los sistemas mainframe a computadoras personales.

Inicialmente las universidades estadounidenses fueron las pioneras en el desarrollo de programas estadísticos tales como el BMD (Biomedical Computer Program) desarrollado por la UCLA en lenguaje Fortran. Cuando las pc's invadieron el mercado, los programas estadísticos más populares fueron desarrollados por los - hasta entonces -, líderes para mainframe, como eran SAS Intitute Inc., SPSS Inc., las Universidades de Pennsylvania y California, que desarrollaban programas como SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences), MINITAB y el mencionado BMD, respectivamente. Cuando éstas pudieron "manejar cálculos estadísticos complejos, los instructores estaban ansiosos por utilizarlas para la enseñanza estadística. La facilidad de tener a cada estudiante trabajando con una máquina interactivamente, representaba un enorme adelanto en términos pedagógicos" HEPC (1993).

¹ Estadística proviene de la palabra ESTADO, pues comenzó siendo parte del aparato de gobierno desde los tiempos en que Salomón y Cesar Augusto realizaron los primeros censos de población

Actualmente existe una gran diversidad de software comercial en el área estadística, lo cual nos obliga como matemáticos a tener una visión general de dichas herramientas, asesorar en la elección de alguna de ellas de acuerdo a las necesidades de los usuarios, además de apoyar a quién lo requiere en su manejo, no olvidando que en muchas ocasiones existen problemas para el usuario (en diversos campos de aplicación) en cuanto a la elección de pruebas a realizar con sus datos.

Este trabajo surge precisamente de la necesidad existente en una de las 14 estaciones meteorológicas del bachillerato de la Universidad Nacional Autónoma de México pues deseaban analizar los datos recopilados en el período 1997-2001, pero no se tenían los conceptos estadísticos formales para realizar esta tarea por parte de los estudiantes que forman el Club de Meteorología que se organiza en cada estación. Esto en lugar de representar un problema, hace que el esfuerzo sea dirigido a explicar de manera didáctica algunos conceptos básicos de estadística, utilizando los archivos del clima de la zona, resultando un material de interés general y con múltiples usos.

En la parte meteorológica este trabajo se ajusta a los lineamientos marcados en la literatura mencionada y en los comentarios recibidos del Dr. Ernesto Jauregui Ostos, del Centro de Ciencias de la Atmósfera (CCA) de la UNAM; con lo que se logró concretar una propuesta para el análisis de los datos recopilados en la estación de la preparatoria número 3. Posiblemente en un futuro cercano, - aún cuando el equipo de la estación no fue creado con fines de investigación- los datos puedan ser utilizados en alguna investigación por parte del CCA, sin que tengan que partir de una base de datos no validados, sino por el contrario con un respaldo de la calidad de ellos.

Y como el coordinador de la estación no solo es responsable de la recolección de los datos climatológicos y de su validación-preservación, sino también de la difusión de su información; los resultados de este informe se pueden compartir mediante la transmisión de los archivos vía correo electrónico o haciendo uso de Internet ya que la estación tiene su propia página de divulgación en la dirección: <http://dgenp.unam.mx/esmet-3/index.html>.

En el primer capítulo se plantea una visión general de los esfuerzos realizados en el análisis estadístico de datos del clima o contaminación en el Distrito Federal y área metropolitana, así como una descripción más detallada del programa de servicio social y los objetivos que plantea cumplir este esfuerzo.

Durante el capítulo 2 hacemos una revisión histórica del software estadístico y sus principales características, resaltando las que son de importancia primordial para la elección de una de estas herramientas, como lo es el programa Statistica que se utilizó a lo largo de este informe.

El capítulo 3 da cuenta de los elementos que conforman el instrumental meteorológico y el software que tiene la estación de la preparatoria 3, también presenta datos que investigamos y que completan la documentación de la estación, aclarando además los conceptos meteorológicos básico.

Ya que el manejo de datos (como es este caso de estudio) involucra la aplicación de métodos estadísticos, se propone en el capítulo 4 un proceso de validación y, posteriormente presentamos alternativas de análisis de estadística descriptiva para concretar algunos informes anuales de las principales variables registradas mediante la estación Davis, sobre todo considerando que son productos de gran utilidad en la investigación climatológica formal y sería una tarea sencilla para los estudiantes, para después proponer usos de la estadística inferencial, especialmente de pruebas de hipótesis con los datos del clima de la zona.

Durante la realización de esta actividad se intentó compaginar y aprovechar las metodologías particulares de la estadística y la climatología, esto trajo consigo un aprendizaje significativo - en ambas áreas- que se puede presentar de igual manera en los miembros del Club de Meteorología de la preparatoria, sobre todo tomando en cuenta que quienes lo forman están motivados, y pueden estarlo más al concebir la magnitud y trascendencia del trabajo que realizan.

**DESCRIPCIÓN
DEL PROGRAMA
DE SERVICIO SOCIAL**

DESCRIPCIÓN DEL PROGRAMA DE SERVICIO SOCIAL



Para vivir es demasiado el tiempo;
para saber no es nada.

Rosario Castellanos

EL LABORATORIO DE ENSEÑANZA-APRENDIZAJE ASISTIDO POR COMPUTADORA EN LA DIRECCIÓN GENERAL DE SERVICIOS DE CÓMPUTO ACADÉMICO

La Dirección General de Servicios de Cómputo Académico (D.G.S.C.A.) presta servicios relacionados con los campos computacionales, a la comunidad universitaria y específicamente a la docente. Desde sus inicios, en 1985, tiene entre sus funciones "establecer sistemas de cómputo académico de las dependencias, vigilando su adecuada operación. Orientar y apoyar a las dependencias tanto en la adquisición y mantenimiento de equipos de cómputo, como para el aprovechamiento óptimo de los recursos institucionales"(WWW,DGSCA). Para realizar lo anterior se organiza en cuatro direcciones:

- 1) Dirección de Telecomunicaciones Digitales
- 2) Dirección de Cómputo para la Docencia
- 3) Dirección de Cómputo para la Administración Académica
- 4) Dirección de Cómputo para la Investigación

Como parte integrante de esta última dirección, se encuentra el Laboratorio de Enseñanza-Aprendizaje Asistido por Computadora (LEAAC) que como tarea primordial tiene la de vincular a los profesores de la Escuela Nacional Preparatoria y Colegio de Ciencias y Humanidades, con el uso de la computadora como herramienta de apoyo educativo.

En el LEAAC se tiene como objetivo ofrecer alternativas de solución a las diferentes necesidades que se presentan en el desarrollo o vinculación con los sistemas de cómputo que apoyan la educación en diversas áreas. Entre algunas de las actividades que realiza el laboratorio se cuentan:

- Desarrollo ó evaluación de software y hardware educativo
- Evaluar, adaptar y asimilar nuevas tecnologías educativas
- Contribuir a la formación de Infraestructura y recursos humanos para el desarrollo de proyectos de enseñanza asistida por computadora
- Establecer criterios y normas para la adecuada aplicación de la tecnología
- Ofrecer estancias a profesores de bachillerato
- Establecer contacto con proveedores de software y hardware educativo
- Fungir como canal de comunicación entre diferentes grupos interesados en el tema
- Organizar cursos y diplomados relacionados con las tecnologías que se analizan en el LEAAC
- Elaborar notas y tutoriales
- Establecer contactos con el extranjero

El LEAAC ha participado en la automatización y modernización de los laboratorios de física, química y biología dentro de los planteles del bachillerato universitario, así como la investigación y programación de software para el mejoramiento de la enseñanza de las actividades experimentales de dichas materias, aunado a la elaboración de la documentación para apoyar estas actividades.

LA ESTACIÓN METEOROLÓGICA Y EL LEAAC

En septiembre de 1996, se reinició la operación de la Estación Meteorológica de la Escuela Nacional Preparatoria plantel 3 "Justo Sierra", que forma parte de un conjunto de estaciones instaladas en los 14 planteles de bachillerato², ubicadas en:
Escuela Nacional Preparatoria (ENP)

- Plantel 1. Xochimilco
- Plantel 2. Iztacalco
- Plantel 3. Eduardo Molina
- Plantel 4. Tacubaya
- Plantel 5. Villa Coapa
- Plantel 6. Coyoacán
- Plantel 7. La Viga
- Plantel 8. Plateros
- Plantel 9. Lindavista

Colegio de Ciencias y Humanidades (CCH)

- Azcapotzalco
- Naucalpan
- Vallejo
- Oriente
- Sur

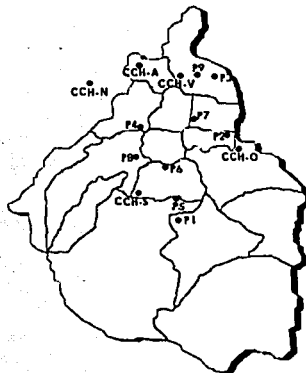


FIGURA 1. Ubicación de las estaciones meteorológicas del bachillerato

² Para mayores detalles sobre el proyecto de estaciones en el bachillerato, consultar ENP-CCH-PIDI (1994)

NOV 21 1997
BIBLIOTECA DE LA UNAM

El proyecto como tal se inauguró en 1992, pero debido a ciertos problemas en cuanto al mantenimiento de los accesorios de la estación meteorológica, a la diversidad de configuraciones del equipo con que se contaba en cada plantel, así como a la diversidad de los perfiles profesionales de los coordinadores de la estación (pues hay geógrafos, físicos, biólogos, etc.), el proyecto se vio retrasado o interrumpido en algunos casos. Por lo que, como se mencionó, a fines de 1996 se inició con nuevo equipo meteorológico automático de la compañía estadounidense Davis Instruments, y con la instalación de equipo de cómputo idéntico en todas las estaciones, salvándose hasta entonces el problema de uniformidad.

De esta manera es que, actualmente las 14 estaciones meteorológicas están operando, pero en docentes de la ENP plantel 3 (Lic. Ligia Kamss, coordinadora de las estaciones de la ENP), ha surgido la inquietud de pasar de la simple recopilación de datos, al análisis, aprovechamiento y conclusiones basadas en dichos datos (como es el delinear las características del microclima de esa zona de la ciudad); para lograr esto, recurrieron al apoyo y asesoría del LEAAC, aprovechando la vinculación que ha existido entre ambos. Partiendo principalmente de que este servicio social había iniciado con el estudio teórico referente a los productos de software estadístico en el mercado y, esta tarea representaría un estudio de caso. Es de esta manera que el trabajo que se expondrá aquí abarcará ambos temas:

- I) Análizar el software estadístico comercial
- II) Un estudio exploratorio de la situación que existe en la estación meteorológica con respecto al análisis estadístico actual y a su vez realizar propuestas para el aprovechamiento de los datos disponibles de los años 1997, 1998, 1999, 2000 y hasta el mes de septiembre de 2001.

1.1 ANTECEDENTES SOBRE ANÁLISIS ESTADÍSTICOS ATMOSFÉRICOS EN EL VALLE DE MÉXICO

Hay que hacer notar que el análisis estadístico de variables atmosféricas metropolitanas, hasta el momento se realiza en parte por el Servicio Meteorológico Nacional (SMN) que cuenta con 72 observatorios meteorológicos distribuidos en todo el país, 65 de los cuales son automáticos y miden los parámetros en forma continua y los transmiten vía satélite al SMN cada 3 hrs; también cuenta con 3500 estaciones climatológicas (miden temperatura, precipitación y viento) en todo el país (alrededor de 54 están en el D.F.) y de igual manera cuenta con 15 estaciones de radlosondeo, 12 radares meteorológicos y una estación terrena para la captación de imágenes de satélite. Es dependiente de la Comisión Nacional del Agua (CNA) y ésta a su vez de la Secretaría del Medio Ambiente, Recursos Naturales y Pesca (SEMARNAP). Cabe mencionar que para el Valle de México solo cuentan con dos estaciones meteorológicas automáticas, una en Tacubaya y otra en el aeropuerto. La CNA maneja por separado otra red pluviográfica.

Otra instancia gubernamental que tiene estaciones que miden variables meteorológicas y que opera desde 1984 es la Red Automática de Monitoreo Atmosférico (RAMA), dedicándose principalmente a la medición de contaminantes,

TESIS CON
FALLA DE ORIGEN

específicamente del Valle de México y se considera de las mejores en el mundo RAMA (1995), actualmente también realiza mediciones de rayos ultravioleta (UV); es el lugar de donde parte la información en casos de contingencias ambientales. La RAMA cuenta con 33 estaciones automáticas distribuidas en el Valle de México (21 en el D.F. y 12 en el Edo. de México), 10 de las cuales tienen la facultad de realizar mediciones de parámetros meteorológicos, aunada a las funciones de detección de contaminantes como ozono, monóxido de carbono, etc. También cuenta con 19 estaciones que conforman la Red Manual de Monitoreo Atmosférico que mide las partículas suspendidas, plomo y sulfatos. Además tiene una unidad móvil de monitoreo, dos estaciones piloto y equipos Rass, ecosonda y de espectrometría de absorción óptica diferencial (DOAS por sus siglas en Inglés). Estas redes forman parte de los Gobiernos del Distrito Federal y del Estado de México. En coordinación con ellos trabaja el Instituto Mexicano del Petróleo, en donde se ha realizado el Estudio Global de la Calidad del Aire, dentro de la Subdirección de Protección Ambiental. También el Gobierno del D.F. cuenta con otra red que parcialmente es pluviográfica.

Por parte de la UNAM se cuenta con varios institutos como el Centro de Ciencias de la Atmósfera (CCA) que tiene más de 20 años de operación y está dividido en cuatro departamentos: teoría del clima, ciencias ambientales, instrumentación meteorológica y meteorología general. En este último existe una sección de Meteorología Urbana dirigida por el Dr. Ernesto Jauregui quien ha realizado importantes investigaciones específicamente sobre el Valle de México como son:

- Efectos de largo período de la urbanización sobre variables del clima
- Posible impacto de la isla de calor sobre las mediciones de la red urbana de tanques de evaporación clase A.
- Efectos de la urbanización sobre la precipitación convectiva
- Valuación bioclimática
- Variaciones espacio-temporales de la humedad
- Balance energético y turbulencia atmosférica

Estos estudios se basan en datos obtenidos de diversas fuentes, como el SMN y la RAMA mencionados anteriormente, así como en observaciones realizadas en coordinación con el Instituto de Investigaciones Eléctricas (IIE), el Departamento de Geografía de la Universidad de Columbia Británica y el Centro de Ecología de la UNAM, con instrumentos del IIE y propios (estación Texcoco y Minería) del CCA.

Las conclusiones de dichos análisis se han presentado en las conferencias:

- 2nd UNAM-CRAY Supercomputing Conference: Numerical simulations in the environmental and earth sciences. Del 21-24 de junio de 1995 en la Cd. de México
- ISEB'95 International Symposium on Environment and Biometeorology. 20-22 de julio 1995. Beijing, China.

Y se han publicado en "Ehecati" que es la revista mensual del departamento de Meteorología General. Cabe comentar que la sección de Física de nubes que es parte del mismo departamento de Meteorología usa la supercomputadora Cray (uno de los principales equipos de cómputo de la UNAM) para implementar un modelo dedicado al estudio de turbulencias en nubes estratocúmulus, este modelo fue desarrollado por el meteorológico británico y se llama Large Eddies Simulation (LES).

También, se realizan estudios meteorológicos nacionales por parte de otros Institutos de la UNAM como son el Instituto de Geografía, en la sección de climatología y, en el Instituto de Ecología se hacen estudios sobre cambio climático. Sin embargo hay que recalcar que en la Universidad solo existe una "red" meteorológica, la del bachillerato y que incluso supera en número de estaciones a la RAMA. Además esta última detecta cuatro variables (velocidad y dirección del viento, temperatura, humedad relativa), mientras que con el equipo con que se cuenta en el bachillerato se detectan las variables anteriores y otras como la presión atmosférica, el punto de rocío y la precipitación pluvial, en periodos que varían de 1 a 120 minutos. Algunas de estas últimas variables tampoco son registradas en las estaciones climatológicas del SMN, pues como se mencionó antes, ahí se registran de manera manual las variables de temperatura, precipitación y viento una vez al día (8:00 a.m.).

Por su parte, los coordinadores de las estaciones meteorológicas del bachillerato realizan un boletín informativo dirigido a su comunidad en donde se incluye información resumida de las condiciones meteorológicas reinantes en la zona durante ciertos lapsos de tiempo (pues varía la periodicidad de esta publicación en cada plantel). Y las estaciones del CCH Sur y preparatorias 3 y 5 difunden su trabajo a través de Internet.

Como se puede apreciar, aunque existen diversas organizaciones involucradas en estudios meteorológicos, los objetivos del análisis de datos difieren, pues en algunos casos se dirigen hacia la Investigación y en otros a la síntesis y difusión de información de interés general.

Este trabajo no se generó a partir de un requerimiento específico de información por parte del usuario, por esto, se realizarán sugerencias para el aprovechamiento -con fines docentes- del material generado por el instrumental meteorológico, mediante un análisis estadístico descriptivo e inferencial (correspondiente a los planes de estudios del CCH y la ENP), y considerando principalmente que hoy en día no es necesario realizar manualmente esta labor.

Para realizar esto se partió de las siguientes ideas:

- ♦ La etapa del bachillerato representa frecuentemente la fase de formación y definición vocacional, por lo que de alguna manera, las actividades que se realizan en la estación meteorológica, pueden influir en la orientación final del estudiante.
- ♦ La asistencia de los estudiantes al club de meteorología es voluntaria y posiblemente motivada por la situación de cambio climático en su entorno.
- ♦ Al revisar los planes de estudios del bachillerato, se observó que solo está integrado curricularmente en forma optativa un curso de estadística lo cual representó de manera inicial una limitante, que puede ser transformada en el reto de mostrar al estudiante (de modo aparentemente incidental) que las matemáticas aplicadas sirven, entre otras muchas cosas, para corroborar los conocimientos adquiridos en otras materias - como la geografía - y generar productos útiles para otras actividades (por ejemplo, la Investigación).

Podemos concluir, haciendo una consideración sobre la trascendencia que estos esfuerzos tienen, pues actualmente la UNAM es miembro fundador de GENie (Global-problematic Education Network Initiative) una iniciativa de la UNESCO surgida en

○ DESCRIPCIÓN DEL PROGRAMA DE SERVICIO SOCIAL

1995 (que busca un contacto estrecho entre especialistas de varias disciplinas y la gente joven), fundamentado ampliamente sobre lo que se conoce como Desarrollo Sostenible y cuyos objetivos son:

- I. Ayudar a los estudiantes a entender la sociedad global - los retos y las oportunidades - con la que se enfrentarán en su vida como adultos.
- II. Capacitar estudiantes para desarrollar las habilidades de pensamiento crítico y solución de problemas para que participen activa y exitosamente en la sociedad del siglo XXI.

El fenómeno del Cambio Climático es un ejemplo claro de fenómenos de Interés global, pues el clima en cualquier punto del planeta depende de las acciones que los individuos realizan incluso de lo que se está dejando de hacer.

De modo que el trabajo que realizan actualmente las estaciones meteorológicas del bachillerato se puede integrar a corto o mediano plazo en un esfuerzo de mayores dimensiones, como lo es GENIE.

1.2 OBJETIVOS Y META

De aquí que el programa de servicio social enfocado para la carrera de Matemáticas Aplicadas y Computación tenga como:

OBJETIVO GENERAL

Usar con fines docentes las estadísticas descriptivas, utilizando los datos generados en la Estación Meteorológica de la Escuela Nacional Preparatoria plantel 3, y a su vez realizar inferencias basadas en la evidencia muestral

OBJETIVO PARTICULAR

Realizar un estudio general del software estadístico y los elementos que lo conforman así como el empleo de alguno como herramienta en el análisis de los datos experimentales, con una orientación didáctica.

OBJETIVO ESPECÍFICO

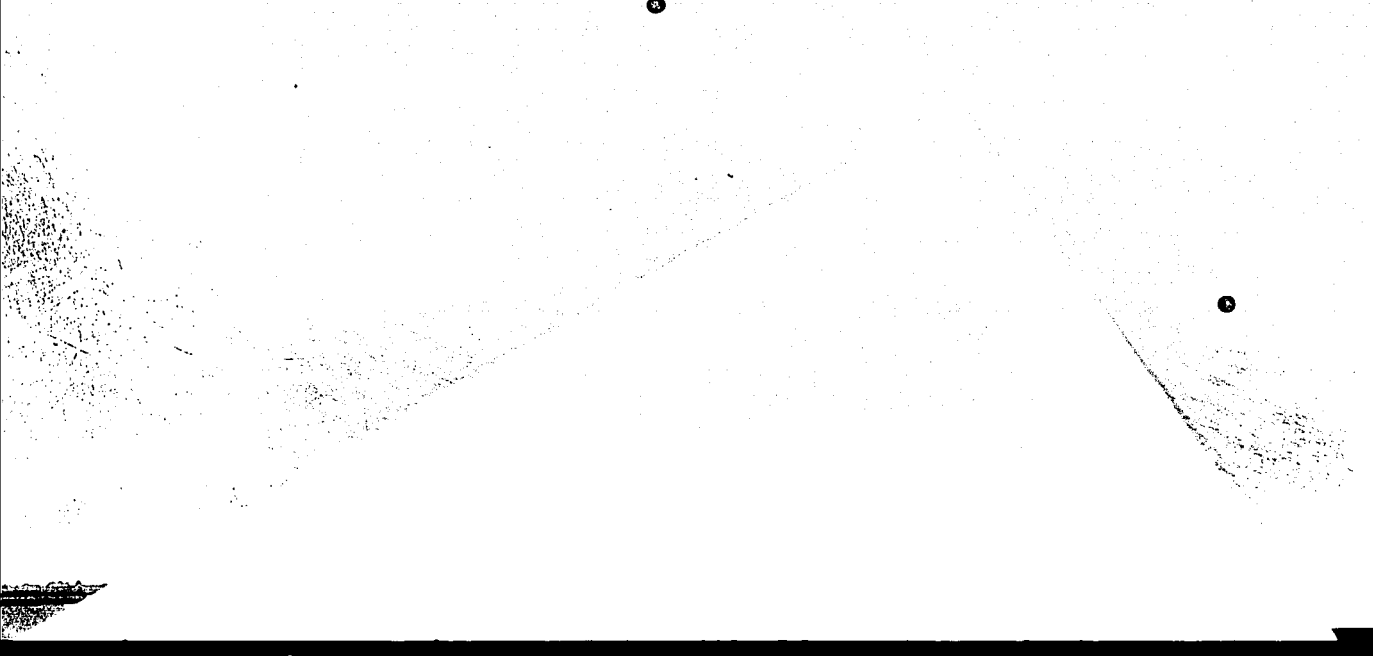
Motivar el interés de los estudiantes del bachillerato, por el estudio de la estadística aplicada, mediante el análisis de los datos climatológicos de la zona geográfica de su escuela.

META

Redactar un manual para que los coordinadores de las estaciones meteorológicas y otros profesores puedan capitalizar estos datos con fines didácticos. Además de ir conformando la información que se planea exponer en una página WEB.

**SOFTWARE ESTADÍSTICO Y
CRITERIOS DE SELECCIÓN**

2



No me explico porqué
fue indispensable que alguien inventara el reloj
y desde entonces todo se atrasa o se adelanta,
la vida se fracciona en horas y minutos
o se quiebra o se para.

Rosario Castellanos

En esta sección se pretende presentar el contexto general en el que se ubica el software estadístico; cómo se ha transformado históricamente; de qué maneras tan variadas se comercializa y cómo está conformado actualmente, para este punto se presentarán - con el objetivo de ejemplificar cada elemento - algunas características del programa que se utilizó a lo largo de la realización del servicio social, dichas características o elementos son los que se consideraron como determinantes al momento de seleccionar un software dedicado a auxiliar en la labor estadística.

2.1 SOFTWARE Y SU CLASIFICACIÓN

Recordemos que de manera muy general se considera que un sistema de cómputo esta integrado por dos elementos: los programas (*software*) y el equipo físico (*hardware*), siendo este último constituido a su vez por memoria, procesador, dispositivos de entrada/salida y discos (dispositivos de almacenamiento). De manera muy sencilla podemos decir que el software es quién le dice al hardware lo que tiene que hacer. De ahora en adelante se utilizarán de manera indistinta los términos software y programas.

Existen diversas clasificaciones del software, algunos autores como Davis (1987) solo consideran dos categorías: software 1) del sistema y 2) de aplicación; pero una de las más completas es la de Sanders (1990), que a continuación se esquematiza:

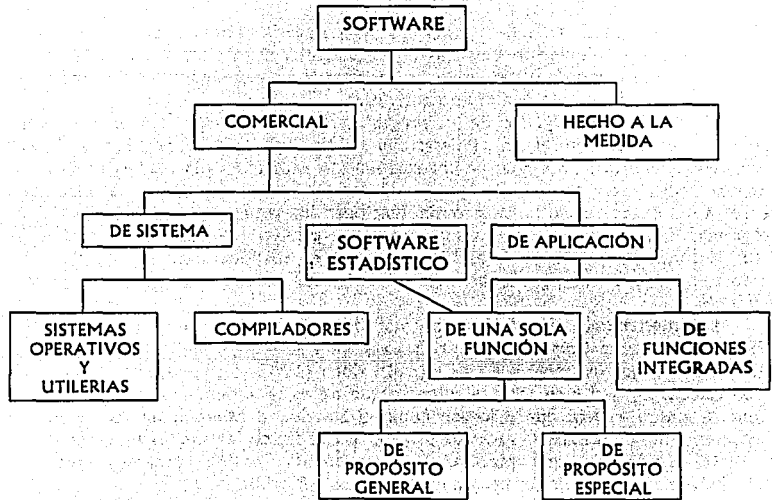


FIGURA 2. Tipos de software

Como se puede apreciar en el diagrama, se inicia por diferenciar dos categorías principales, el software hecho a la medida y el comercial; el primero es creado principalmente para cubrir los vacíos que el segundo pueda dejar y de esta manera satisfacer necesidades específicas de una organización o persona.

El software comercial -"generalmente se ajusta a las necesidades de muchos usuarios"- está subdividido en dos: de sistema, en el cual se incluye como ejemplos muy claros el sistema operativo, los programas traductores (compiladores), también se pueden incluir los programas dedicados a protocolos de comunicación, (Davis considera a los manejadores de base de datos dentro de esta categoría). Y respecto a los programas de aplicación es más exhaustivo dividiendo a su vez esta categoría en:

Software de funciones integradas, que "incluye todas o casi todas las siguientes funciones - procesamiento de palabras, manipulación de hojas de cálculo, manejo de archivos o bases de datos, envío de comunicaciones, preparación de gráficas y creación de índices"

Software de una sola función

- De propósito especial.- como son los juegos y los programas dirigidos a una clase de usuarios determinada como áreas de recursos humanos, floristas, bufetes legales, farmacias, etc. En el campo de la estadística existen algunos programas que se dedican a un solo tipo de análisis como las series de tiempo, análisis de supervivencia, etc.
- De propósito general.- realizan una tarea específica como el procesamiento de textos, manipulación de hojas de cálculo, etc. pero pueden ser utilizados en cualquier tipo de oficina u hogar, soportando las necesidades de un grupo más grande de usuario que el tipo anterior.

Si consideramos algunos de los programas comerciales más conocidos veremos que las actuales tendencias en el desarrollo de sistemas se han inclinado a diseñar programas que inicialmente eran de una sola función pero que cada vez incorporan más funciones adicionales, haciendo confusa la ubicación de un programa en las categorías de software de una sola función o de funciones integrales; éste es el caso de los programas estadísticos, pues no solo permiten el estudio estadístico o el análisis de datos, sino que por la propia necesidad de este campo adiciona un manejo gráfico muy potente, aunado al manejo de archivos con diversos formatos lo cual es altamente requerido.

La International Association for Statistical Computing³ publicó hasta el año 1997, en su revista *Computational Statistics & Data Analysis*, una guía anual sobre el software estadístico existente a escala mundial Koch (1996 y 1997), clasificándolo en varias categorías, y si bien incluye software que no es totalmente estadístico como el Maple, e ignora a otros que sí lo son como el Statview o Minitab; da una referencia muy interesante de la extensa gama de programas estadísticos existentes actualmente.

La clasificación que manejan en esta revista, podría tomarse como una subdivisión más extensa de los programas de una sola función.

³ Una sección de The International Statistical Institute.

2.2 LA ESTADÍSTICA Y LA NECESIDAD DE SU UTILIZACIÓN EN DIVERSAS ÁREAS DE CONOCIMIENTO

La propia naturaleza de la Estadística allienta la creación y utilización de programas de cómputo dedicados a esta ciencia como un auxiliar en sus procedimientos de investigación. Aclaremos que estadística es mucho más que un nombre para describir los hechos o figuras de los deportes o la economía. Es una rama de las matemáticas que es parte integral de la investigación científica y elemento determinante para la toma de decisiones, así lo citan varios autores:

Al analizar el papel de la estadística en la ciencia Craddock (1968) consideró que el mayor logro de la primera, es decir, los métodos Inferenciales, formaban parte de los fundamentos de todas las ciencias; para algunas de ellas son la herramienta utilizada en el trabajo de investigación y en otras están tan profundamente arraigados que incluso se utilizan como parte de la metodología particular.

Para David Hunstberger (1983), la estadística "se relaciona con el desarrollo y aplicación de métodos y técnicas para obtener, analizar e interpretar datos cuantitativos de tal manera que la confiabilidad de las conclusiones basadas en los datos puedan evaluarse objetivamente por medio de proposiciones de probabilidad". Para él, el propósito de esta disciplina es apoyar con sus métodos a quienes se dedican a la investigación en todas las ciencias cuantitativas, aunque los problemas puedan ser muy diversos "la estadística puede considerarse como la tecnología del método científico".

Mientras que George Canavos (1988) la define como "el estudio de los fenómenos aleatorios" cuyo aspecto primordial es "la obtención de conclusiones basadas en los datos experimentales. Este proceso se conoce como inferencia estadística", la cuál es inductiva ya que va de lo particular (muestra) a lo general (población).

También indica que hay cuatro elementos que caracterizan a los problemas estadísticos:

1. La población de interés y el procedimiento científico que se empleó para muestrear la población
2. La muestra y el análisis matemático de su información
3. Las inferencias estadísticas que resulten del análisis de la muestra
4. La probabilidad de que las inferencias sean correctas

De manera similar Ya-Lun Chou (1972) menciona las etapas de toda investigación estadística:

- I. Definir lo que va a ser investigado. Ya que "las técnicas estadísticas, por muy refinadas y precisas que sean, no pueden ayudar a alcanzar decisiones si son aplicadas a datos inapropiados"
- II. Diseño de la muestra, es decir, el diseño del experimento
- III. Compilación de datos. Se refiere a los métodos usados para obtener información pertinente a la muestra
- IV. Organización y descripción de datos. Cálculo de estadísticas descriptivas.

V. Decisión o Inferencia final. "Trata de responder, basándose en las estadísticas de la muestra al problema o la pregunta original formulada, que siempre se refiere a los parámetros de la población".

El punto 2 y IV que mencionan Canavos y Chou respectivamente, como parte de la metodología particular de la estadística, involucran un trabajo de análisis, generalmente sobre un gran número de datos, lo cual fundamenta la necesidad de apoyarse en el software estadístico como herramienta para aligerar y agilizar el trabajo, en cualquier área de conocimiento en que se apliquen sus métodos o procedimientos, como es la psicología, la Ingeniería, la economía, la medicina, la biología, las áreas bioquímicas, meteorológicas, astronómicas, etc.

Un artículo de la revista PC Magazine [Yacono(1994)] señalaba en su título que "Cualquiera puede ser un estadístico", refiriéndose a la facilidad con la que actualmente se puede aprender a usar el software de este tipo, pero aún cuando tenga esa cualidad, además de su gran utilidad, no puede pensarse que pueda sustituir al Intelecto de quien lo utiliza.

Vale la pena conocer el desarrollo histórico de este tipo de software ya que nos permitirá comprender sus características actuales.

2.3 BREVE HISTORIA DEL SOFTWARE ESTADÍSTICO

Es importante conocer como se han ido transformando los programas estadísticos, principalmente como respuesta a la necesidad de adecuarse a las plataformas de hardware que a través del tiempo se han presentado.

Veremos que los primeros programas estadísticos surgieron dentro de una plataforma para mainframe, después estos programas se tuvieron que migrar a las minicomputadoras y posteriormente se adaptaron a las computadoras personales, pero también notaremos cómo los desarrollos de programación han tomado ventaja de los avances de hardware para no tan solo hacer portable sus aplicaciones sino también mejorarlas en bien de sus usuarios.

2.3.1 LOS PROGRAMAS ESTADÍSTICOS PIONEROS

El primer paquete que logró un amplio reconocimiento fue BMD (Biomedical Computer Program), el cual fue desarrollado en la Universidad de California en Los Angeles (UCLA). Era una librería de programas en Fortran diseñada para manejar una variedad de funciones estadísticas usando una estructura de datos compartida. Como un primer esfuerzo, fue muy conocido entre la comunidad científica y académica, pero tenía algunas dificultades para trabajar, tal como el uso limitado que permitían sus instrucciones. El primer manual de usuario fue editado en 1963 por W.J. Dixon [Moore(1988)].

Otro programa de la primera generación, SAS (Statistical Analysis System), fue desarrollado en Carolina del Norte por programadores del estado y universitarios de Duke, ellos son: Anthony J. Barr, Goodnight, Sall y Helwing; y cuyo fin era analizar datos agrícolas. El primer manual de usuario fue creado en 1972, sin embargo se

realizó una revisión y corrección, y se puede considerar su lanzamiento formal en 1976. Fue un logro de muchos estadísticos y programadores que continuamente estaban "reinventando la rueda" HEPC(1993): este iba creciendo con nuevos algoritmos para manejar operaciones en particular, pero no se tenía un formato estándar con el cual compartir estos esfuerzos. Una asociación no lucrativa, el Instituto SAS, fue creada para formar los inicios de la estandarización y desarrollo del SAS. El grupo, que más tarde se privatizó, desarrolló una versión comercial del programa, desarrollado únicamente para el mainframe IBM 360. Esto les permitió crear un sistema manejador de base de datos extremadamente flexible, y el SAS se convirtió rápidamente en el líder en el campo del cómputo estadístico, aunque de alguna manera el hecho de ser un sistema propietario no permitía que fuera portable.

El tercer producto inicial de estadística fue SPSS (Statistical Package for the Social Sciences). Desarrollado en la Universidad de Chicago (1970) por Norman H. Nie, Hadlai Hull, Jenkins y Steinbrenner y por parte de la Universidad de Alberta participó Dale Bent. Originalmente se había destinado para usarse en cuantificaciones de supervivencia y otros datos de las ciencias sociales, pero a causa de su relativa flexibilidad y al hecho de que podía correr en una variedad de computadoras, también fue ampliamente usado como una herramienta de enseñanza. El prefacio del segundo manual para el usuario (1975) indicaba que en la primera distribución del producto se realizaron 60 instalaciones y para esa segunda versión del producto se tenían ya 600 instalaciones del producto.

Un rasgo que compartían los tres paquetes estadísticos originales era su falta de interactividad. Reconociendo que esto era menos que óptimo para la enseñanza a los estudiantes sobre la estadística, tres instructores de la Universidad del estado de Pennsylvania -Thomas Ryan, Barbara Ryan, y Brian Joiner- se reunieron a principios de los setentas para desarrollar un paquete estadístico que fuera más apropiado para la enseñanza.

El resultado fue un programa para mainframe llamada Minitab. Era interactivo, permitiendo que los estudiantes usaran sus terminales para ver los procedimientos aplicados a sus datos, pero esta flexibilidad tenía un precio: solo se podían manejar conjuntos de datos limitados. Aún así, debido a su interactividad, logró una buena acogida en el campo de la educación.

Los principales lenguajes en que se desarrollaron estos primeros programas para mainframe, así como para minicomputadoras fueron los siguientes: APL, BASIC, C, Fortran, Pascal y PL/I; de ellos el que más prevaleció fue Fortran y esto es comprensible, ya que con mayor disponibilidad se encontraba en las computadoras mainframe. Las principales implementaciones se realizaron en Fortran ANSI 1966 o en el estándar 1977, aprovechando sus características de procesamiento numérico tales como el manejo de arreglos multidimensionales, aritmética compleja y subrutinas de compilación independiente.

BASIC no fue muy utilizado ya que no permitía una estandarización simple, ni la modularidad. Varios sistemas de excelente calidad fueron desarrollados en lenguaje APL como el Statgraphics y Statpack.

A su vez estos programas pioneros desarrollaron posteriormente su propio lenguaje de programación dedicado a sus respectivos usuarios, cada uno de los paquetes estadísticos mencionados (BMD, SAS, SPSS, MINITAB) asignaron a su correspondiente lenguaje de usuario el mismo nombre del programa.

2.3.2 LA GENERACIÓN PARA PC

Hay ocasiones en que los avances tecnológicos originan un profundo cambio en la sociedad y sobre disciplinas específicas. La introducción de los microprocesadores desató ese tipo de cambios en el uso de las técnicas de análisis estadístico, pues brindó a muchas personas capacidades sofisticadas de cálculo estadístico a bajo costo.

Sabemos que los primeros usos de las microcomputadoras o computadoras personales fueron en la industria, la escuela y las casas. Los primeros programas eran pequeñas aplicaciones de negocios, programas educativos y juegos, por mencionar algunos. Es claro que el crecimiento de la popularidad de las PC's ha estado acompañado del diseño de software para usarse en ellas y viceversa (como ejemplo tenemos a Microsoft).

Cuando las memorias de las computadoras personales empezaron a ser más baratas y la velocidad de los cálculos fue mejorada, los programas para análisis estadístico comenzaron a aparecer. En 1980 solo un puñado de dichos programas se vendían [Woodward (1982)], y para 1985, Alan C. Elliott de la Universidad de Texas menciona que existen alrededor de 250 programas estadísticos, pero hace hincapié en que no todos se pueden considerar "paquetes", ya que algunos de ellos solo se dedicaban a un tipo de análisis (eran de propósito especial) ya sea econométrico, biomédico, sociológico, etc.

No es sorprendente, que los primeros desarrollos de paquetes estadísticos para computadoras personales los realizaron los líderes en mainframe, quienes migraron sus aplicaciones a una nueva plataforma. Los resultados fueron la generación de paquetes estadísticos para PC's manejados mediante un diálogo en línea de comandos, los cuales, en general, eran difíciles de usar y podían manejar únicamente un conjunto de datos de tamaño limitado. Sin embargo tenían la ventaja de que eran muy familiares para quien los hubiera trabajado en plataforma mainframe.

Se considera que existió toda una "explosión" de dichos programas estadísticos para microcomputadoras, desarrollados muchos de ellos por programadores que no tenían la experiencia de las grandes compañías creadoras de los paquetes para mainframe. Dichos productos no contaban con un análisis por parte de la comunidad estadística que avalara su calidad, incluso en los años 80's se detectaron por parte de la American Statistical Association algunos programas que tenían errores de exactitud muy evidentes [Elliott (1985)]; este factor, al igual que el uso indiscriminado e inadecuado de estos programas originó que algunas investigaciones incluyeran resultados incorrectos, lo que obviamente invalidaba toda conclusión obtenida. De todas maneras no hay que sorprenderse de que el principal foco de usuarios de los programas estadísticos, no sean los propios matemáticos o especialistas en estadística, sino los investigadores de otras ciencias como economía, genética, medicina, etc.

Respecto al primer problema mencionado, continuamente en las universidades, Institutos, asociaciones y sociedades a nivel mundial, se realizan Investigaciones especializadas en el cómputo estadístico y publican tanto en medios impresos como electrónicos sus conclusiones, de manera que esto representa implícitamente un control y una evaluación periódica de los programas estadísticos existentes, tanto en su contenido matemático, como en el diseño o los algoritmos. Respecto al segundo punto se insiste en la continua capacitación de los usuarios de estos programas o en la búsqueda de asesoría especializada.

Algunos de los problemas iniciales que fueron superados durante esta generación de programas fueron: el de precisión, haciendo uso de adelantos como el co-procesador matemático, acompañado de avances en los lenguajes de programación; los tiempos de procesamiento se vieron disminuidos al mejorar la velocidad de reloj de los procesadores; así como los adelantos en las tecnologías de almacenamiento y memoria.

Por su parte, los creadores de software estadístico han sometido a un fuerte análisis a sus algoritmos, y han hecho uso de las capacidades gráficas que se les dieron en esta plataforma, para beneficio de los usuarios.

2.3.3 SOFTWARE ESTADÍSTICO HOY

La Introducción del cómputo basado en ventanas y el uso de Interfaces gráficas (ver sección 2.4.2.1 *Diálogo*) ofrecieron mayor potencial para diferentes programas estadísticos (usan menús o iconos, en lugar de líneas de comandos).

Además, las nuevas habilidades gráficas de las computadoras abrieron el camino para los programas estadísticos que proporcionaban capacidades de despliegues avanzados, usándose tanto para la instrucción como para la investigación. El énfasis que se puso en las gráficas y en la manipulación directa de datos ofreció un ambiente que habilitó al usuario una interacción personal con sus datos, manipulándolos y visualizándolos en la pantalla de la computadora en tiempo real.

Se ha visto que como otros programas de cómputo, el software estadístico se ha transformado notablemente. Hace solo cinco años, el mercado de estos productos estaba poblado de programas de difícil aprendizaje y uso, así como considerablemente lentos. Sin embargo, esto ha cambiado, actualmente a la par de los avances de hardware (como la familia de procesadores Pentium), de la sencillez de las interfaces Windows y acompañadas de mejoras en los propios paquetes estadísticos, se pone a disposición de los usuarios programas de una potencia extraordinaria "capaces de afrontar análisis estadísticos hasta hace un tiempo reservados a grandes plataformas de cálculo" Herrán (1996).

Hace 10 años, el hecho de realizar un análisis de regresión con 10 000 casos tomaba medio minuto, ahora se obtiene en pocos segundos. Sin embargo, el principal logro de los actuales programas estadísticos para PC, no es la potencia, sino su facilidad de uso. Los actuales usuarios de paquetes estadísticos basados en Windows 95/98 y 2000, requieren dedicar muy pocas horas para aprender cálculos avanzados.

Los programas en plataforma Windows, no están elaborados bajo un mismo molde, existen programas con diseños *integrales*, los cuales tiene un gran conjunto de análisis en el mismo programa; pero también existen tendencias marcadas de realizar programas con estructura *modular*, conformados por uno o más módulos principales y otros más que se pueden ir adicionando de acuerdo a las necesidades del usuario, esto hace aparentar que algunos programas de propósito general estén integrados por varios de propósito específico. Lo cual provoca que exista una gran diversidad de precios. Sobre este punto explica Lucas (1983):

Las empresas fabricantes buscan paquetes de uso general para incrementar su potencial de mercado y reducir la necesidad de modificar el paquete. Para lograr este propósito pueden seguirse dos estrategias: un paquete puede producirse de manera que presente ciertos parámetros de entrada o tablas; el cliente proporciona muchos de estos parámetros solamente una vez al iniciar el paquete. Otro enfoque es desarrollar un paquete con varios módulos; en este caso el usuario configura un paquete que contiene solo los módulos que a él le son necesarios. Muchas veces las empresas que desarrollan software combinan los dos enfoques.

Durante la realización de esta práctica profesional, se pudo conocer y constatar - gracias a que los universitarios contamos con excelentes recursos humanos (académicos e investigadores), material bibliográfico con literatura especializada en una infinidad de temas, así como la infraestructura de cómputo y comunicaciones para el acceso a Internet y correo electrónico-, que en las más de tres décadas de existencia formal de los programas o paquetes estadísticos para computadora, todos estos se han transformado radicalmente, lo que nos confirma que en el área de cómputo, las investigaciones, estudios o discusiones sobre cualquiera de sus elementos, ya sea hardware, software, servicios o telecomunicaciones, pasan en muy poco tiempo a formar parte de la historia.

Prueba de ello es que para 1985 la American Statistical Association consideraba la existencia de más de 200 programas estadísticos [Elliot (1985)], muchos de los cuales ya no existen actualmente.

En las Guías de Software estadístico más recientes [Koch (1997)] se consideran generalmente alrededor de 90 programas de este tipo, la mayoría de ellos orientados a su uso en microcomputadoras, sin embargo, también se mencionan programas para otras plataformas menos populares. Estas listas o guías generalmente los clasifican en programas de propósito general, de propósito específico, bibliotecas, *shareware* y *teachware/learnware*, sin embargo, no siempre son completas ya que algunos de los programas pioneros como el Minitab no se incluyen, aún cuando son los principales "sobrevivientes" a los cambios de plataformas de hardware y software.

Conjuntando los dos elementos mencionados anteriormente: la pronta pérdida de vigencia de los programas de cómputo (específicamente los estadísticos, que son los de nuestro interés), así como el gran número de ellos, parecería ilógico intentar estudiar cada uno de los programas y además evaluarlos o compararlos, con el fin de lograr una selección correcta; sobre todo considerando que durante la investigación se verificó que ya existen trabajos (publicaciones periódicas o páginas de Internet) encaminados a dicha

labor*. Las referencias que se han mencionado hasta ahora y las que a continuación se dan, pueden servir para mostrar los esfuerzos realizados en cuanto a comparación o evaluación de software estadístico:

- ♦ Lock, Robin H. "A comparison of five student versions of statistics packages". *The American Statistician*. Vol. 47, No. 2, p. 136-145. Mayo 1993.
- ♦ Lurie, Phillip M. "A review of five statistical packages for Windows". *The American Statistician*. Vol. 49, No. 1, p. 99-107. Febrero 1995.

Y respecto a las creaciones recientes de listas de software de este tipo podemos consultar en:

- ♦ Koch, Armin y Haag, Uwe. "The statistical software guide '95/96". *Computational statistics & data analysis*. Vol. 21, No. 2, p. 231-250. Febrero 1996.
- ♦ Koch, Armin y Haag, Uwe. "The statistical software guide '96/97". *Computational statistics & data analysis*. Vol. 24, No. 1, p. 107-131. Marzo 1997.

Y en las páginas de:

- ♦ Indiana University (Stat/Math Center) (EUA)
<http://www.indiana.edu/~statmath/stat/index.html> (abril 2002)
- ♦ The american statistician association. Section on statistical computing.
<http://www.statcomputing.org> (abril 2002)
- ♦ Statistical solution
<http://www.statolusa.com> (abril 2002)
- ♦ ProGAMMA bv.
<http://www.gamma.rug.nl/> (abril 2002)

De hecho, la mayoría de las publicaciones que se han mencionado, no reconocen a ningún programa por encima de otro, pero como es de esperarse cada productor de software estadístico proclama a su producto como el mejor, lo cuál obviamente responde a técnicas mercadotécnicas. Y de hecho, realizar comparaciones solo tiene sentido si se realizan con un propósito determinado, como encontrar el máximo costo/beneficio, para alguna aplicación en particular, etc.

Aún al margen de lo comentado, se puede elegir correctamente un programa estadístico haciendo algunas consideraciones generales.

* Aunque al iniciar este trabajo se pretendía realizar evaluación y comparación de software estadístico, gracias a que se realizó una búsqueda bibliográfica con el fin de conocer y ampliar el conocimiento sobre los aportes anteriores, se pudo evitar duplicar esfuerzos en una labor ya lograda.

2.4 CARACTERÍSTICAS DETERMINANTES EN LA ELECCIÓN DE UN PROGRAMA ESTADÍSTICO

Está claro que en muchas ciencias y actividades se requiere el análisis estadístico de datos y que el software estadístico es una herramienta útil para lograr este fin. Si debemos elegir entre las muchas posibilidades existentes debemos tener en mente que lo más importante es responder: *¿Qué necesidades de información se tiene y quién va a utilizar el programa?*, así como considerar los elementos de hardware y software con que se cuenta.

Ubicando los requerimientos del usuario, se puede elegir entre un programa de propósito general u otro de propósito específico. Revisando el contenido estadístico y gráfico de algunos programas de propósito general como SPSS, MINITAB, SYSTAT, SAS, SIGMASTAT, STATGRAPHICS y el propio STATISTICA; se verificó que es muy similar y aún habiendo muchas variantes en cuanto a su diseño u organización (sobre todo en cuanto a los programas modulares), en general incluyen la totalidad de elementos mencionados en la siguiente sección.

Los programas de propósito específico, cubren un subconjunto del universo de posibilidades existentes dentro del software de propósito general, sin embargo, aquellos son más exhaustivos, pues incluyen pruebas que no se encuentran en estos últimos. De aquí la necesidad de reconocer las necesidades de análisis que tenga el usuario.

2.4.1 CONTENIDO ESTADÍSTICO Y GRÁFICO DE LOS PROGRAMAS DE PROPÓSITO GENERAL

Enseguida se enumeran los temas que son abordados por un software estadístico de propósito general, señalamos con un asterisco (*) los tópicos que coinciden con un texto de estadística para nivel licenciatura. En lo que respecta a los temas que se incluyen en la currícula del bachillerato (CCH y ENP) son cubiertos en el punto señalado como estadística básica, pues comúnmente se incluye el cálculo de medidas de tendencia central, desviación, prueba t, correlación y regresión y la observación de la distribución normal y binomial.

Para un examen detallado de los diferentes programas estadísticos se recomienda consultar las páginas de Internet que aparecen en las referencias ya que periódicamente tienen actualizaciones.

La parte gráfica de los programas de este tipo son una herramienta que enriquece su uso, aplicación y aprendizaje.

- ☞ Estadística descriptiva (Básica) *
- ☞ Análisis de Regresión *
- ☞ Análisis de Varianza (ANOVA/MANOVA) *
- ☞ Análisis de Cluster *
- ☞ Análisis Factorial *
- ☞ Análisis Discriminante *
- ☞ Análisis Canónico *
- ☞ Tablas de Contingencia *
- ☞ Tablas de Referencia cruzada *
- ☞ Análisis de correspondencia *
- ☞ Control de procesos *
- ☞ Diseño Experimental *
- ☞ Análisis de series de tiempo *
- ☞ Pruebas no paramétricas *
- ☞ Análisis de supervivencia *
- ☞ Análisis de confiabilidad *
- ☞ Simulación y distribución *
- ☞ Estimación no lineal *
- ☞ Análisis log-lineal *
- ☞ Componentes de varianza *
- ☞ Escala multidimensional *
- ☞ Modelamiento estructural de ecuaciones *

GRÁFICAS*

- ☞ Gráficas en 2-D:
 - Histogramas
 - Diagramas de cajas
 - Gráficas lineales
 - Gráficas de puntos
 - Gráficas de barras
 - Gráficas de pie
 - Distribuciones de probabilidad
 - Gráficas de dispersión
- ☞ Gráficos en 3 dimensiones (3-D)
 - Lineales
 - De puntos
 - De barras
 - De bloques
 - De listones
 - De espigas
 - Gráficas de contorno
 - Gráficas de superficie

- ☞ Gráficos de control de calidad (Pareto)
- ☞ Matrices de dispersión
- ☞ Opciones de edición de gráficas
- ☞ Zoom

2.4.2. INTERFACE

Respondiendo a la pregunta *¿Quién va a utilizar el software estadístico?*, podemos apoyarnos en la forma que Davis (1987) clasifica a los usuarios en clases que no son excluyentes:

- a) Desarrolladores y no desarrolladores
- b) Novatos o expertos
- c) Ocasionales o habituales
- d) Principales o secundarios

El mismo autor señala que lo principal es lo que se refiere a la experiencia del usuario que consiste en dos componentes:

- ◆ Conocimiento sintáctico, es decir, el que se adquiere con la frecuente utilización de un sistema en particular
- ◆ Conocimiento semántico, el manejo general que tiene el usuario de los conceptos de los sistemas de cómputo

Es importante utilizar estos criterios ya que valiéndose de la experiencia de un usuario, se puede elegir una interface similar a las conocidas anteriormente, es decir, que exista consistencia entre ellas.

Uno de los elementos determinantes para considerar que un software es amigable, es decir, fácil de aprender y de usar, y que a su vez, redunde en un uso efectivo del sistema, es el diseño de la interface. Esta "es una frontera compartida" que proporciona la base para la comunicación entre el usuario y el sistema de cómputo. Veamos algunos elementos que conforman el diseño de una interface.

2.4.2.1. DIÁLOGO

De manera general se considera que existen cinco tipos de operación de interface: mediante uso de comandos, por menú, de tipo gráfica, tipo formulario (cajas de diálogo) y con lenguaje natural.

- Comandos. - es la más compleja ya que el usuario debe aprender un cierto "lenguaje" para operar el programa. Los comandos tienen un formato preespecificado y están acompañados de uno o más argumentos para especificar los detalles. Ejemplos de este tipo de diálogo son: MS-DOS, SQL, el lenguaje propio de los programas estadísticos.

- Menú. - lleva al usuario por una serie de procedimientos para captura o procesamiento de datos, dando o pidiendo información en cada nivel del programa, pues se parte de en un menú principal que contiene otros niveles (submenús) en los cuales las operaciones se van haciendo más detalladas. Es fácil de comprender y en consecuencia se requiere menor tiempo para su aprendizaje que el uso de comandos. Para los usuarios expertos no es muy cómodo su uso. Actualmente la mayoría de los menús son de tipo "desplegables", es decir, se exhibe todo el tiempo un renglón con las opciones que tiene y al elegir una de ellas, se despliega a detalle el siguiente nivel de opciones. También existen los menús "short-cut" que incluyen las opciones más frecuentemente utilizadas en algún programa y se activan casi siempre con el botón derecho del mouse.
- Gráfica. - muestra las opciones a elegir en forma de dibujos o símbolos gráficos, llamados iconos, que se pueden escoger utilizando un dispositivo señalador como el Mouse o el lápiz electrónico. Este tipo de diálogo combina el uso de cajas de diálogo, iconos, menús y menús desplegados simulando un escritorio, lo cual facilita la interacción humano/máquina. Ejemplo: Windows.
- Formulario o caja de diálogo. - utiliza la filosofía de "llenar espacios", usa el posicionamiento automático del cursor para captar los datos de entrada y las solicitudes del usuario. También puede presentar varias opciones entre las que el usuario debe marcar la que desea ejecutar.
- Lenguaje natural. - trata de lograr que el usuario se comunique con la computadora mediante un lenguaje hablado como el español o el inglés. Estas interfaces son competencia de la inteligencia artificial y de alguna manera hacen innecesaria la existencia de los diálogos explicados antes. Sin embargo actualmente existen algunos programas con un uso y "vocabulario" limitado. Ejemplo: software de reconocimiento de voz de IBM, ViaVoice.

En el caso específico de *Statística*, debido a su nacimiento en plataforma Windows, permite que sea totalmente consistente con ella y utiliza obviamente un diálogo de tipo gráfico principalmente y eventualmente con las cajas de diálogo.

2.4.2.2. DISEÑO DE PANTALLAS

Debe ser claro, para evitar confusiones y libre de información irrelevante, sin embargo debe incluir todos los datos que se encuentren relacionados lógicamente con una misma tarea. Las interfaces gráficas hacen uso de gráficos y colores que mejoran la claridad del despliegue de pantallas, así como de la división de ventanas que permiten un manejo simultáneo de varios elementos, como puede ser en el caso de un software estadístico el manejo de gráficas, datos y reporte escrito.

En el caso particular de *Statística*, se manejan cuatro tipos de pantallas

- 1) Para el manejo de los datos de entrada (*Spreadsheet*), que es parecido a una hoja de cálculo, sin embargo, por la terminología propia de la estadística cambian los conceptos; por ejemplo, en lugar de hablar de renglones y columnas, se manejan los

- términos de casos y variables, respectivamente. Los archivos de datos tienen un formato *.sta.
- 2) Para desplegar los resultados estadísticos después de un análisis, se utiliza una hoja con la misma presentación que la anterior, llamada *Scrollsheet*, pero con algunas variaciones en cuanto a las funciones que se pueden realizar con ella. Las hojas de resultado tienen un formato *.scr y pueden ser salvadas como "datos" para ser analizadas nuevamente.
 - 3) La que nos permite desplegar y editar los elementos que conforman el análisis gráfico de los datos
 - 4) También podemos almacenar los resultados en un archivo de texto, que puede ser revisado o editado simultáneamente a la operación.
- Se pueden consultar en el ANEXO A, las facilidades que tiene la barra de herramientas de *Statistica*, correspondientes a cada tipo de pantallas que se han mencionado.

2.4.2.3. RETROALIMENTACIÓN Y AYUDA

Otra parte fundamental y determinante en la elección de un software- cualquiera que sea -, es el sistema de retroinformación y ayuda, que se puede presentar de varias formas:

- Informando de alguna manera, al usuario, que recibió su solicitud. Generalmente, se satisface la necesidad de retroinformación del usuario, cuando en pocos segundos el sistema da una respuesta; o si la tarea requiere emplear más tiempo se puede desplegar un mensaje o una barra de avance del proceso.
- Mediante ayuda adicional haciendo uso de mensajes, códigos de error o más frecuentemente utilizando documentación y tutores en línea.

Otro elemento que forma parte de la interface y que esta ligado al sistema del que estamos hablando, es el TIEMPO DE RESPUESTA, es decir, el tiempo que transcurre entre una solicitud al sistema por parte del usuario y el inicio de la presentación de la respuesta.

Statistica cuenta con manual impreso y electrónico, así como un "consultor" para guiar al usuario sobre qué tipo de prueba es conveniente realizar con sus datos. Y mediante su página Web se tiene acceso a un amplio libro electrónico y glosario de términos estadísticos (en inglés).

Además, cuando una tarea requiere emplear mayor tiempo de respuesta que el de una actividad sencilla, se despliega en la parte inferior de la pantalla, como parte de la barra de status, la barra de avance del proceso, para retroinformación al usuario.

2.4.2.4. CONTROL DE ERRORES

Una interface amigable y fácil de usar, integra en su diálogo la característica de minimizar los errores que el ser humano puede cometer al operar el sistema, o sea, que tiene un adecuado control de errores.

Dicho control se puede dar en cuatro dimensiones:

- Prevención. Proporcionando explicaciones específicas al usuario, de como realizar una tarea.
- Detección. Esto se refiere a que el sistema debe identificar clara y explícitamente el error cometido.
- Corrección. Debe ser directa y que solo se requiera el tecleo de los datos erróneos.
- Recuperación. Se refiere a la capacidad de deshacer algo que se ha realizado previamente.

Respecto a este punto, *Statistica* trabaja correctamente en cuanto a las tres primeras dimensiones para el control de errores, ya que al utilizar frecuentemente cajas de diálogo, previene que el usuario pueda introducir datos erróneos y en caso de que esto ocurra, lo detecta rápidamente e informa al usuario mediante un mensaje cómo corregir y así continuar la operación.

En cuanto a la recuperación, la opción deshacer no existía en las primeras versiones del programa *Statistica*, lo cual era un inconveniente al iniciarse en su uso. Las últimas versiones del software ya corrigen este problema.

2.4.3. PLATAFORMA

Entre los elementos que influyen de manera determinante en la elección de un programa estadístico se encuentra la plataforma de cómputo con que se cuente, incluso en algunos casos es una limitante. Respecto a los programas estadísticos de propósito general que se encuentran en el mercado, principalmente están orientados hacia su uso en Pc's, cosa que parece obvia al corroborar las ventas de equipo de cómputo a nivel nacional y mundial. Sin embargo existen versiones de varios programas estadísticos que pueden ser instalados en mainframe. Y respondiendo a las tendencias actuales del mercado los productores se inclinan a desarrollar sus últimas versiones hacia Windows e incluso Mac y no hacia DOS como se hacía anteriormente.

Como se mencionó previamente, en la estación meteorológica del bachillerato se cuenta con una computadora personal con procesador 486 y se trabaja bajo Windows 3.11 (posteriormente se instaló otra PC Pentium); *Statistica* fue diseñado desde su nacimiento como un programa orientado para trabajar con Windows - aunque actualmente tiene versiones para Mac-, de manera que respecto a este punto, la instalación del programa no representa ningún problema. Statsoft Inc. menciona que el espacio en disco duro ocupado por el programa completo es de 12 MB (tampoco es esto un impedimento en el caso de la estación meteorológica). Incluso si hubiera problema de espacio en disco duro, se podría realizar la instalación mínima, que aún así, cubre con las necesidades básicas de análisis y manejo de datos en la estación.

2.4.4. CARACTERÍSTICAS GENERALES

En este punto nos referimos a algunos elementos que no son netamente estadísticos, pero que influyen en la operación del software estadístico y en su rendimiento y que por supuesto también son indicadores importantes para la elección del programa. Por ejemplo:

- ◆ Velocidad y precisión
- ◆ Las capacidades de manejo de datos como son:
 - Número máximo de casos y variables soportados
 - Opciones de importación y exportación de datos
 - Posibilidad de ordenar, verificar y explorar datos
 - Facilidad en la edición de datos
 - Tratamiento de datos perdidos
 - Vínculos DDE, OLE y DLL
 - Manipulación de matrices
- ◆ Facilidad de elaboración y calidad de presentación de reportes y gráficas
- ◆ Lenguaje de programación para el usuario y macros
- ◆ Soporte técnico y capacitación en el uso del programa

El programa *Statistica* no tiene limitación específica sobre el número de casos y variables que puede manejar, excepto por las propias limitaciones de memoria del equipo de cómputo y diversas opciones de manipulación de datos. En cuanto a las posibilidades para importar y exportar datos estas pueden verse en la Figura M 3 (en la sección *MANUAL*), adicionalmente hay diferentes opciones para el manejo de bases de datos y la interface estándar ODBC. Tiene un lenguaje similar al de los sistemas de 4G (*SCL - Statistica Command Language*) para facilitar su operación estadística y otro lenguaje de programación para transformación de datos (*MML - Megafle Manager Language*). Cuenta con procedimientos para la optimización de velocidad y manejo de memoria. Y usa innovaciones computacionales de bajo nivel en los programas para lograr una "cuádruple precisión". Y se integra perfectamente a Windows mediante los vínculos DDE, OLE y DLL. En el 2002 Statsoft tiene ya un producto llamado *WebStatistica* para aplicaciones que involucran Internet y también la facilidad de programar macros con Visual Basic que es uno de los lenguajes orientados a objetos mas conocido.

2.4.5. COSTO

Indiscutiblemente el dinero que requerimos emplear para la adquisición de un software, cualquiera que sea, es elemento determinante para la decisión final.

Cuando tenemos claro lo que el usuario necesita de un software estadístico y existen ya algunas opciones para elegir, tenemos que decidir entre los productos que nos den mejor rendimiento a menor costo; ya que de un mismo producto generalmente se han desarrollado diversas presentaciones y de esto depende el costo. Por ejemplo,

actualmente los productos comerciales más conocidos que cubren las funciones estadísticas de uso general y para análisis gráfico básico tienen un costo muy variable, el intervalo de los precios encontrados en varios programas⁵ hasta el primer trimestre de 2002 eran entre 700 y 1600 dólares (licencia anual), correspondiente al módulo base y de estadísticas básicas de algunos productos que se venden de manera modular o los equivalentes programas que se venden de manera integral.

Específicamente sobre *Statistica*, han existido varias presentaciones, incluso esto se modifica constantemente pues anteriormente se vendía una versión completa por el precio de \$1180 + IVA y la versión reducida *Quick Statistica* costaba \$680 + IVA (fines de 1999); a mediados de 1998 se puso a la venta un módulo de redes neuronales y un poco después se vendía la versión para uso industrial (*Statistica Industrial System*) que requería alguno de los productos anteriores, mientras que durante 1999 se pusieron a la venta otros módulos adicionales orientados principalmente a la industria. En el 2002 el software se vende en forma modular, costando aproximadamente \$1060 + IVA el módulo base (completiendo fuertemente con el módulo base de SPSS con valor de \$1062 + IVA), y se puede adecuar a las necesidades del usuario con módulos adicionales que varían entre 555 y 665 dólares. El módulo base incluye estadísticas básicas, estadística no paramétrica, análisis de varianza, regresión lineal y métodos de regresión múltiple.

Aún cuando el software estadístico no es ampliamente demandado (considero que esto se debe principalmente a la falta de conocimientos en la materia), es importante para quienes como nosotros los matemáticos u otros especialistas en diversas áreas de conocimiento requerimos en algún momento fundamentar mediante inferencias las conclusiones obtenidas en estudios realizados a lo largo de nuestra vida profesional. Y utilizarlo en el ambiente laboral como herramienta en la toma de decisiones.

Como hemos visto, desde su surgimiento ha existido una gran diversidad de software estadístico y ahora hay que considerar que mediante Internet es posible tener acceso a casi cualquier distribuidor de software estadístico en el mundo, por lo cual no representa gran problema obtener información sobre ellos; pero cuando se requiere soporte o capacitación sí se debe considerar la disponibilidad local del producto que se desea utilizar, para evitar obstáculos en el aprovechamiento del programa.

En general, en los elementos citados están englobados los puntos determinantes para la elección de un programa que auxilie en la labor estadística. En todos los casos se mencionó al software *Statistica* que se utilizó en este trabajo resaltando en cada punto desarrollado, las características que encontramos en dicho programa estadístico, sin embargo debemos señalar que las razones principales que nos llevaron a seleccionarlos son:

- 1) Que las Instancias con quienes se puede llegar a tener Intercambio de datos como lo son Institutos de la UNAM específicamente el Centro de Ciencias de la Atmósfera y, el Instituto Mexicano de Petróleo manejan este software lo que implica un ahorro de tiempo en cuanto la importación y exportación de datos.

⁵ Como SPSS, SAS, MINITAB, Shazam, NCSS, *Statistica*, etc.

- 2) El hecho de que los usuarios no tuvieron ningún problema para adaptarse al software ya que es totalmente consistente a la Interfaz gráfica de Windows y que es ampliamente conocida. Aquí existían varias opciones pues los principales programas cuentan con versiones disponibles para ambiente Windows.
- 3) El punto más importante fue la facilidad de realizar la validación de datos, que por su número no era factible de realizarse tan rápida y fácilmente con otros programas como por ejemplo con Excel o Statview.
- 4) El desembolso por el software fue nulo, pues se disponía de este y otros programas como Statview, SigmaStat, SPSS, Systat y una versión para DOS de Minitab, sin contar con las versiones de prueba que proporcionan los creadores de estos y otros productos mediante su página en Internet, los cuales no pudieron utilizarse con el total de datos pues el número de casos y variables que manejan son limitados.

Por último, un artículo publicado en Internet durante 1998 (Kuonen, 1998) sobre el tema, menciona siete puntos clave para considerar y seleccionar un software estadístico moderno (algunas de estas características coinciden con las señaladas por este trabajo):

- 1) Desempeño, refiriéndose a la velocidad y al manejo de memoria.
- 2) Que sea escalable y que explore todo el poder de cómputo disponible.
- 3) Que sea predecible el tiempo que tomará la ejecución de una tarea.
- 4) Compatibilidad, se refiere a que el código en que se escriba un paquete estadístico debe estar libre de las barreras que impone el uso de un solo sistema operativo, el autor propone a Java como opción de desarrollo.
- 5) Amigable con el usuario. Señala que en general debería contener un buen diseño de Interface, intuitiva y que simplifique el uso de nuevas tecnologías.
- 6) Agentes inteligentes. Que se haga uso de sistemas expertos para asistir y advertir al usuario en los errores más comunes.
- 7) Buena presentación tanto en los datos como en los resultados, pues un buen trabajo pierde calidad sin este elemento.

El autor compara nueve paquetes estadísticos y concluye que actualmente en el mercado del software estadístico "la mejor elección" no existe y deben ser consideradas cuidadosamente las necesidades específicas del usuario. Además que es tiempo de escribir programas estadísticos que permitan "explorar las posibilidades del nuevo milenio". Como inicio de esta propuesta se podría considerar el desarrollo en Internet de herramientas como el WEBSTAT Y RWEB (WWW,GASP) que son herramientas muy útiles para el análisis numérico y que se encuentran desarrolladas directamente para utilizarse en línea y están abiertos a cualquiera que tenga acceso a Internet y sin ningún costo, además que se pueden utilizar datos propios o puede accederse a las bases de datos que están integradas a dichas páginas. Seguramente el uso masivo de Internet

revolucionará y transformará el mercado del software estadístico como lo ha venido haciendo en muchas otras áreas*.

* Se recomienda leer el artículo "Statistical tools on the world wide web" *The American Statistician*, Vol.32 No5 Agosto 1998 (<http://www.amstat.org/publications/tas/west.html>)

**DESCRIPCIÓN
DE LAS
ESTACIONES METEOROLÓGICAS
Y
LOS DATOS EXPERIMENTALES**

3

El que tiene un porqué en la vida,
siempre encuentra el cómo

Nietzsche

3.1 ANTECEDENTES

En la UNAM existen catorce planteles de enseñanza media superior, nueve pertenecen a la Escuela Nacional Preparatoria y cinco al Colegio de Ciencias y Humanidades. En cada uno de ellos, se tiene instalada una estación meteorológica que es operada por estudiantes de cualquier grado o semestre del bachillerato y que integran un Club de meteorología, asesorados por un profesor del plantel, quien funge como coordinador de dicha estación.

Este "Proyecto de Diseño, Construcción, Instalación y Operación Estudiantil de Estaciones Meteorológicas en el Bachillerato", surgió por un acuerdo entre los directores de los planteles y el entonces rector de la UNAM, Dr. José Sarukán Kermez, en 1991. El objetivo principal del proyecto es interesar y concientizar a los alumnos en cuestiones ambientales, despertar su interés en la investigación e inquietud por contribuir a la solución de los problemas ambientales del Valle de México, además de incorporar conocimientos básicos de meteorología a su formación integral.

Desde entonces se trabajó de manera conjunta entre:

1. El Programa de Integración de Docencia e Investigación (PIDI), como coordinador del proyecto.
2. La Escuela Nacional Preparatoria (ENP) y el Colegio de Ciencias y Humanidades (CCH), comprometiéndose a proporcionar las facilidades de espacio físico, acondicionamiento de dicho lugar, equipo de cómputo y la asignación de un académico responsable del proyecto en cada plantel.
3. El Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS), quién diseñó, construyó e instaló el primer equipo meteorológico con que contaron las estaciones; también realizó la capacitación de los coordinadores.
4. La Dirección General de Asuntos del Personal Académico (DGAPA), quién cubrió el costo del proyecto y administró los recursos.

El 28 de septiembre de 1992, se inauguró en la ENP 8 el Programa de Estaciones Meteorológicas del Bachillerato Universitario y se dio inicio al curso de coordinadores; mientras que durante 1993 se instalaron los equipos en cada uno de los catorce planteles; con esta estación automatizada se registraban, mediante un mástil de sensores, tres variables: temperatura, velocidad y dirección del viento, medidos respectivamente por un termómetro, un anemómetro y una veleta. Estos datos eran almacenados y procesados mediante un software llamado "Meteoro", instalado en equipos de cómputo muy variados, con procesadores que iban desde 8086 hasta 80486.

Debido a que los sensores de la estación meteorológica estaban instalados en la azotea y no contaban con abrigo, aunado a la falta de mantenimiento, así como a la

falta de homogeneidad en los equipos de cómputo utilizados en ese momento; fue necesario cambiar por completo dichos elementos. Actualmente se tiene instalado en todas las escuelas del bachillerato, un mismo equipo meteorológico y computadoras (algunas conectadas a RED-UNAM), con procesador 486; que fueron instalados a finales de 1996.

En el catálogo 1995 de "Instrumentos Meteorológicos de Precisión" elaborado por Davis Instruments, aparecen varios de los modelos fabricados por esta industria, el más completo de ellos, ya que combina las características de los demás es el Weather Monitor II, este equipo es el que actualmente está operando en cada plantel.

En el siguiente cuadro se describen las variables meteorológicas que se registran actualmente en la estación meteorológica del plantel 3 de la ENP "Justo Sierra", y en que sistema de unidades se ha establecido trabajarlas (ya que esto puede variar, pues el equipo permite establecer por ejemplo para medir la velocidad del viento, kph, mph, nudos o m/s).

TABLA 1. Variables que se registran mediante Weather Monitor II

VARIABLE	UNIDADES DE MEDICION
FECHA	DD-MM-AA
HORA	FORMATO DE 24 HORAS
TEMPERATURA INTERIOR	GRADOS CENTÍGRADOS
TEMPERATURA EXTERIOR	GRADOS CENTÍGRADOS
TEMPERATURA MÁXIMA EN EL PERÍODO	GRADOS CENTÍGRADOS
TEMPERATURA MÍNIMA EN EL PERÍODO	GRADOS CENTÍGRADOS
PRESIÓN ATMOSFÉRICA	MILIBARES
HUMEDAD RELATIVA INTERIOR	%
HUMEDAD RELATIVA EXTERIOR	%
PUNTO DE ROCÍO	GRADOS CENTÍGRADOS
VELOCIDAD DEL VIENTO	M/S
RACHA (VELOCIDAD MAX. DEL VTO)	M/S
DIRECCIÓN DEL VIENTO	16 RUMBOS
FACTOR DE ENFRIAMIENTO POR VTO.	GRADOS CENTÍGRADOS
LLUVIA	MILÍMETROS

3.2 ALGUNOS CONCEPTOS BÁSICOS DE METEOROLOGÍA

El estado del tiempo siempre es objeto de nuestra atención y en ocasiones de la plática cotidiana; algunas veces nos hace sentir cansados, molestos, irritados; otras optimistas, reconfortados pero jamás dejamos de considerarlo. "El conocimiento del clima ha sido desde siempre una inquietud para el hombre ya que sus actividades de abrigo, sustento y recreo dependen en gran parte de ellas" Mercado (1993).

La "materia prima" de toda investigación meteorológica (ya sea con fines hidrológicos, agrometeorológicos, para evaluar la calidad del aire, etc.), son las observaciones meteorológicas las cuales llevan de manera implícita la evaluación o medición de uno o varios elementos climáticos, esto se realiza con el fin de obtener información sobre el estado momentáneo de la atmósfera, para ser plasmados en forma sinóptica y que se use como la base de los pronósticos climáticos diarios.

Un elemento climático se define como cualquiera de las propiedades o condiciones de la atmósfera las cuales, en conjunto, describen el estado físico del tiempo o el clima en un lugar dado, para un determinado momento o período de tiempo. Cada elemento meteorológico observado puede también ser considerado como un elemento climático. De estos elementos los más comunes son: temperatura, humedad, precipitación, presión atmosférica y movimiento de la atmósfera (expresado por medio de la velocidad y dirección del viento). Otro grupo de elementos climáticos consiste en aquellos fenómenos meteorológicos que ocurren en el seno de la atmósfera o en la superficie de la tierra, tales como nubes, niebla, neblina y otras obstrucciones a la visibilidad, evaporación, duración de la insolación radiante, así como la temperatura, la humedad y el viento en la atmósfera libre.

Las observaciones meteorológicas son realizadas en las estaciones meteorológicas, cuyo objetivo específico es proporcionar información referente al clima correspondiente al lugar donde se efectúa la observación. El resultado de las observaciones difiere de otros lugares, de mes en mes y de año en año. Las estaciones son clasificadas, de acuerdo a la Organización Meteorológica Mundial (OMM) de la siguiente manera:

- 1) Por su finalidad:
 - a) sinópticas
 - b) climatológicas
 - c) agrícolas
 - d) especiales
 - e) aeronáuticas
 - f) satelitales
- 2) De acuerdo a la magnitud de sus observaciones:
 - a) principales
 - b) ordinarias
 - c) auxiliares
 - d) adicionales

* Gómez, S. Benjamín y Arteaga, Ramón. *Elementos básicos para el manejo de instrumental meteorológico*. Ed. Continental. 2ª Impresión. México 1988.

3) Por el nivel de observación:

- a) superficie
- b) altitud

4) Por el lugar de observación:

- a) terrestres
- b) aéreas
- c) marítimas

De acuerdo a lo anterior, la estación meteorológica de la ENP plantel No.3 "Justo Sierra", así como las demás instaladas en planteles del bachillerato de la UNAM son: climatológicas, ordinarias de superficie y terrestres. Sin embargo, desde el punto de vista más estricto no se puede considerar que actualmente conformen una red meteorológica, ya que a pesar de que son del mismo tipo no son administradas de manera conjunta y no existe interrelación entre ellas, ni existe un sitio en donde se concentre la información generada en cada nodo. En todas ellas el instrumental realiza y registra las observaciones automáticamente.

Para lograr que los datos que se van a recopilar sean confiables, la OMM recomienda que al instalar una estación meteorológica se reúnan algunas condiciones, estas son en general: un proceso de planificación y diseño (ubicación y equipamiento) y considerar los factores climáticos que habitualmente influyen sobre el clima: latitud, altitud, topografía, etc. Es importante que las observaciones realizadas sean hechas bajo condiciones normalizadas, en horas fijas y de acuerdo con las prácticas establecidas tanto para la observación como para la exposición de los instrumentos ya que estos datos pueden ser útiles para propósitos de aplicación o investigación.

Los archivos que conforman una base de datos climatológica pueden tener muchas aplicaciones y generalmente están a la disposición de quien los requiera, pero de nada servirían si no se encuentran cuidadosamente puestos al día y son conservados. "Los registros climatológicos de la región aunque de solo 100 días, proveen una información cuyo valor es miles de veces el costo de obtenerlos tienen valor en proporción a la frecuencia e inteligencia de los usos que se les den".

En los reglamentos técnicos de la Organización Meteorológica Mundial (OMM) se menciona que "para que los archivos adquieran auténtico valor, deberán contener datos climatológicos de garantía durante un período de tiempo de 30 años aproximadamente" OMM(1979) pues se considera que con una serie de datos más corta no se puede asegurar que estén incluidos ejemplos de todos los diferentes fenómenos meteorológicos observados en proporciones correctas y con el mismo intervalo de variación como para representar auténticamente el clima del lugar.

Incluso para comparar los climas de distintos lugares se utilizan dos tipos de medias:

1) Promedio de un período: es la media de una serie continua de observaciones de un elemento climático de por lo menos 10 años.

¹ OMM-SMN-SARH. *Compendio de apuntes para la formación del personal meteorológico. Clase III. Climatología.* México 1980.

2) Normal climatológica: es la media de una serie continua de observaciones de un elemento climático durante un período específico de 30 años.

En este trabajo no se deja de considerar lo anterior, pero también menciona la OMM, que cualquier documentación climática es útil desde el momento que se registra y se archiva. "Así, mientras que las series largas son necesarias para describir la totalidad de un clima, las series más cortas - de unos cinco o seis años- pueden dar respuesta a cientos de preguntas importantes y pueden servir de valiosa guía para conocer el tiempo que se espera en un futuro inmediato, después del período de tiempo cubierto por la predicción diaria." OMM(1979). De tal manera que nos dedicaremos a preparar normales provisionales que sean la base de una larga serie, como las que se requieren en el trabajo meteorológico formal, partiendo de datos anuales o en las series que se puedan recolectar a lo largo de un año escolar, que igualmente son de importancia climatológica.

Otro de los lineamientos de la OMM señala que en lo posible toda información estadística basada en datos climatológicos, debe acompañarse de información referentes a la localización geográfica y elevación de las estaciones meteorológicas, así como la homogeneidad de los datos y los períodos establecidos para la observación, por lo que enseguida se documentan estos puntos.

3.3 DOCUMENTACIÓN DE LA ESTACIÓN

NOMBRE: Escuela Nacional Preparatoria Plante No. 3 "Justo Sierra"
DIRECCIÓN: Av. Eduardo Molina 1522 casi esq. Av. San Juan de Aragón.
COLONIA: Salvador Díaz Mirón
DELEGACIÓN: Gustavo A. Madero.
COORDINADORA : Lic. En Geografía. Ligla Kamss

Los siguientes datos corresponden a las poblaciones que se encuentra más cercanas a la ENP 3^a.

ASPECTOS GEOGRÁFICOS

Se considera Zona Urbana.

Localidades Principales más cercanas

	LATITUD NORTE		LONGITUD OESTE		ALTITUD
	Grados	Minutos	Grados	Minutos	
San Juan de Aragón	19	28	99	04	2240
Guadalupe Tepeyac	19	28	99	07	2240

Elevaciones Principales más cercanas:

	LATITUD NORTE		LONGITUD OESTE		ALTITUD
	Grados	Minutos	Grados	Minutos	
Cerro El Guerrero	19	30	99	06	2440
Cerro Gachupines	19	30	99	06	2440
Cerro Tepeyac	19	30	99	06	2440

Fisiografía:

Provincia: Eje neovolcánico. *Subprovincia:* Lagos y Volcanes de Anáhuac

Sistema de Topoformas: llanura lacustre.

Geología:

Era: cenozoica *Periodo:* cuaternario

Tipo de roca: sedimentaria

Unidad litológica: lacustre

Clima: semiseco templado

Regiones, cuencas y subcuencas hidrológicas:

Región: Pánuco *Cuenca:* R. Moctezuma *Subcuenca:* L. Texcoco-Zumpango

Corrientes de Agua: Canal de desagüe, Los Remedios y Consulado(entubados).

Cuerpos de Agua: Lago San Juan de Aragón (artificial).

Horario de Observaciones: 24 hrs. No se había establecido de manera fija. Se inició con períodos de 15 minutos pero en ocasiones se intercalaba con otros de 30/ 60 minutos.

Para detectar los elementos climatológicos citados en la Tabla I, se requieren sensores específicos con los que se encuentra equipada la estación meteorológica, a continuación se presenta una breve descripción y definición de ellos.

3.4 DESCRIPCIÓN DE LOS INSTRUMENTOS METEOROLÓGICOS

Weather Monitor II

Este instrumento permite visualizar en su pantalla, la velocidad del viento y su dirección en la rosa de los vientos, además una flecha indica si la presión atmosférica está bajando o subiendo. Mediante las teclas de función se pueden establecer la fecha, hora y presión barométrica o atmosférica, después de haber instalado el equipo por primera vez. Y con las teclas de temperatura y humedad se puede alternar la indicación de estas en cuanto al interior o exterior. También se pueden fijar alarmas. Si así se establece pueden visualizarse secuencialmente todas estas variables en unidades métricas o inglesas. Sus dimensiones son 13.3 x 14.9 x 7.6 cm.



FIGURA 3. *Weather Monitor II*

Recordemos algunos conceptos de meteorología referentes a las variables mencionadas hasta el momento:

A) La temperatura es el elemento climático más utilizado como indicador de tendencias y variabilidad del clima, además de ser un factor determinante en los modelos de simulación climática, esto convierte al termómetro en uno de los instrumentos meteorológicos más utilizados. El modelo instalado en la estación permite utilizar para su registro grados centígrados o Fahrenheit con una precisión de $\pm 0.5^\circ \text{C}$.

B) La presión atmosférica en una superficie horizontal es la fuerza por unidad de área ejercida por el peso de la atmósfera. La presión en una estación meteorológica es igual al peso de una columna vertical de aire que va desde la superficie terrestre hasta los límites superiores de la atmósfera. El barómetro permite al observador determinar los cambios en la presión atmosférica y a su vez, esta información se utiliza para definir la trayectoria de tormentas y verificar modelos de simulación climática. La presión

• DESCRIPCIÓN DE LAS ESTACIONES METEOROLÓGICAS Y LOS DATOS EXPERIMENTALES

atmosférica al aire libre y bajo techo es la misma. Los cambios en esta variable y la dirección del viento, medidos simultáneamente, nos brindan los mejores indicios para anticipar las alteraciones que puede experimentar la situación climática local. El Weather Monitor II mide la presión atmosférica en pulgadas (") ó milímetros(mm) de mercurio, ó milibares (Mb); detecta valores con una precisión de ± 1.7 Mb.

C) La humedad relativa es el contenido de agua en la atmósfera, expresado como un porcentaje que va de 0 a 100%. La humedad del aire se puede medir con instrumentos manuales o automáticos. El instrumento manual más común es el psicrómetro. Para las estaciones del bachillerato de la UNAM, la humedad interior es la especificación del equipo meteorológico se registra con una precisión de ± 1.7 Mb.

D) Las variables referentes a la velocidad y dirección del viento son de gran importancia en el estudio de los contaminantes en el ambiente. El Weather Monitor II Incluye un anemómetro. El anemómetro mide la velocidad y dirección del viento.

atmosférica al aire libre y bajo techo es la misma. Los cambios en esta variable y la dirección del viento, medidos simultáneamente, nos brindan los mejores indicios para anticipar las alteraciones que puede experimentar la situación climática local. El Weather Monitor II mide la presión atmosférica en pulgadas (") ó milímetros(mm) de mercurio, ó milibares (Mb); detecta valores con una precisión de ± 1.7 Mb.

C) La humedad relativa es el contenido de agua en la atmósfera, expresado como un porcentaje que va de 0 a 100%. La humedad del aire se puede medir con instrumentos manuales o automáticos. El instrumento manual más común es el psicrómetro. Para las estaciones del bachillerato de la UNAM, la humedad interior (por especificación del equipo meteorológico) se registra con una precisión de $\pm 5\%$.

D) Las variables referentes a la velocidad y dirección del viento son de suma importancia en el estudio de los contaminantes en el Valle de México. Para medir esta variable se utiliza un anemómetro. El creado por Davis Instrument para el modelo Weather Monitor II incluye los sensores de velocidad (anemómetro de cazoletas) y dirección del viento (veleta). Sus componentes resisten vientos huracanados, al tiempo que son sensibles a las brisas más ligeras. La dirección del viento se mide en grados (arcos) ó en la rosa de los vientos se especifican 16 rumbos. Y la velocidad se puede determinar en mph, kph, nudos o m/s, esta última unidad es la que se utiliza en las estaciones del SMN, la RAMA y las estaciones del bachillerato de la UNAM, en esta se logran detectar vientos con velocidad de hasta 78.2 m/s y precisión de $\pm 5\%$.

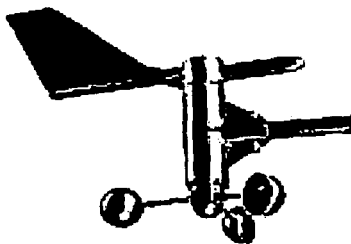
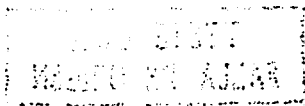


FIGURA 4. Weather Monitor II. Veleta y anemómetro

Los accesorios opcionales del Weather Monitor II, que también fueron adquiridos para las estaciones del bachillerato son un sensor exterior de temperatura/humedad, un pluviómetro y el software Weatherlink, que se describirán enseguida.



Sensor exterior de temperatura/humedad

Permite medir la humedad relativa y el punto de rocío del exterior⁹, además de la temperatura. Incluye 12 mts. de cable. El rango de temperatura exterior que detecta la estación equipada con este equipo va de -45°C a 60°C , con una precisión de $\pm 0.5^{\circ}\text{C}$. En cuanto al punto de rocío el rango va de -99°F a 140°F , o lo que es igual -73°C a 60°C con una precisión de $\pm 2^{\circ}\text{C}$. La precisión con que se registra la humedad es de $\pm 5\%$.



FIGURA 5. *Weather Monitor II. Sensor exterior*

Colector de lluvia - Pluviómetro de balancín

Mide la lluvia diaria y acumulada. Esta construido cuidadosamente para cumplir con las pautas de la OMM, tiene un diseño de balde autovaciante e inclinable que es excepcionalmente exacto. Requiere mínimo mantenimiento, ya que esta elaborado de materiales muy resistentes y es fácil de conectar. Mide de 6.5" a 8.7" de diámetro x 9.5" de altura. Incluye cable que se puede extender.

La precipitación de lluvia diaria se puede medir en pulgadas o milímetros, se encuentra en el rango de 0.0 a 999.8 mm y la lluvia acumulada va del rango de 0.0 a 9999 mm. Con una precisión de $\pm 2\%$



FIGURA 6. *Weather Monitor II. Pluviómetro de balancín*

⁹ El punto de rocío es la temperatura a la cuál el aire alcanza la saturación, es decir, se condensa (deficiencia tomada de la página del SMN).

3.5 DESCRIPCIÓN DEL SOFTWARE DE LA ESTACIÓN - WEATHERLINK

En el sitio destinado para ocupar la estación meteorológica, se tiene instalada una computadora personal dedicada al almacenamiento de la información climatológica, para un posterior análisis y presentación gráfica. Para lograr esto se conecta el cable weatherlink que une al Weather Monitor II con el puerto serial de la computadora. Tiene la opción de efectuar observaciones en una estación remota, para lo cual se dispone de un adaptador para módem compatible con Hayes.

El software tiene los siguientes requerimientos:

- Equipos compatibles con IBM, XT, AT, PS/2, etc. con 512K de memoria convencional libre
- Monitor y tarjeta gráfica tipo Hercules, CGA, EGA, VGA o compatible
- Sistema operativo MS-DOS o PC-DOS 2.1 o superior
- Un puerto serie

Soporta los puertos serie RS-232 números 1,2,3 o 4 y puede realizar impresiones en equipos láser o matriciales.

Davis Instrument proporciona:

- a) Los adaptadores serie de 9 o 25 patillas
- b) 2.5 m de cable
- c) Manual de instrucciones detallado
- d) Disco flexible con el software, 5 ¼ o 3 ½

Este programa (en la versión que se tiene en la estación) está orientado principalmente para crear gráficos; representa instantáneamente en una sola pantalla el boletín meteorológico (Figura 7); permite elegir una variable para ser representada gráficamente en una base diaria, semanal, mensual o anual; puede representar simultáneamente dos días, semanas, meses o años con referencia a una variable especificada; realiza la comparación de dos variables (como máximo) en un mismo gráfico y también se pueden comparar los datos originados en distintas estaciones meteorológicas (cada una con su propio Weatherlink), sin embargo, no se pueden generar reportes con los datos, como los que se necesitan para un análisis más detallado, ni tampoco detecta si existen datos erróneos en su base de datos.

"El weatherlink almacena los datos hasta que son transferidos" a la computadora, en períodos a elegir entre 1, 5, 15, 30, 60 o 120 minutos, "y transfere cada 1, 5, 15, 30, 60 o 120 días"¹⁰. Se puede respaldar la información en disco flexible o disco duro. Se puede exportar los datos a hojas de cálculo como Excel, Lotus 1-2-3, o a Dbase.

¹⁰ Davis Instruments. *Instrumentos meteorológicos de precisión. Catálogo 1995.*

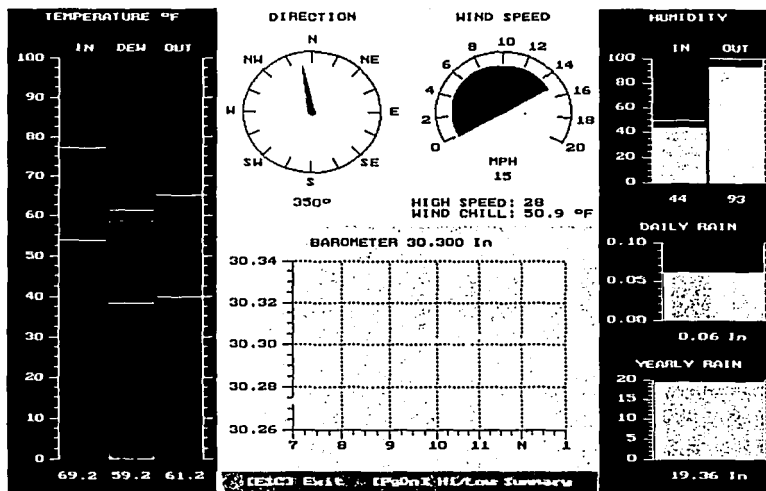


FIGURA 7. Boletín meteorológico de Weatherlink

3.6 ACTUAL MANEJO ESTADÍSTICO DE LOS DATOS

Al asistir las primeras veces a la estación meteorológica de la ENP No.3 "Justo Sierra", fue posible conocer el equipo meteorológico descrito anteriormente (Weather Monitor II) y, qué variables eran detectadas por él, se confirmó en qué unidades se registraban y se exportó la muestra de un mes en formato .TXT, para comenzar a trabajar con los datos, pero principalmente para saber si podían ser migrados al software con el que se contaba y con el cuál se trabajó, sin tener que recapturar o modificar la base de datos de manera significativa. También se verificó que las estadísticas presentadas en la revista "Atmósfera 3", que publica trimestralmente el Club de Meteorología, son generadas en su mayoría por el boletín creado por el software propio de la estación (Weatherlink), y por otra parte se realiza manualmente un Informe mensual que resume el comportamiento día por día de: la temperatura exterior en una base bitoraria, el registro de la velocidad máxima del viento en dicho día, así como las temperaturas máxima y mínima.

● DESCRIPCIÓN DE LAS ESTACIONES METEOROLÓGICAS Y LOS DATOS EXPERIMENTALES

Ejemplo del formato usado para recopilar los datos:

TABLA II. Hoja de registro estadístico usado en la estación de la ENP plantel 3

INFORME BIHORARIO

FECHA: _____

	1	2	3	30	31
HORA						
02:00						
04:00						
06:00						
:						
22:00						
24:00						
VEL/VTO						
TEM MAX						
TEM MIN						

También se ha realizado, con la información de un mes específico el análisis de gráficas (elaboradas manualmente) para detectar la manera en que se ve afectada la temperatura exterior con la entrada de un frente frío, conforme lo informa el Servicio Meteorológico Nacional.

Como se puede ver, había (y hay) una enorme posibilidad de trabajos estadísticos por realizar sobre ese cúmulo de datos.

ANÁLISIS ESTADÍSTICO

ANÁLISIS ESTADÍSTICO

4

A nadie le faltan fuerzas,
lo que a muchísimos les falta es voluntad

Victor Hugo

En esta sección se verá el proceso tanto de exportación como de importación de datos hacia el software estadístico y las pruebas de validación a las que se debe someter una base de datos climatológica. También la forma en que se puede organizar la información (como lo sugiere la estadística, mediante tablas de frecuencia e histogramas) y la obtención de las estadísticas de la muestra que en formas muy variadas son gran parte del trabajo básico en la climatología (estadística descriptiva), iniciando por un análisis univariado y posteriormente bivarlado; y por último una prueba de hipótesis (estadística inferencial); todo lo anterior dándole un enfoque docente. Se podrá notar que ante la necesidad y el deseo de organizar y presentar la parte descriptiva, se plantea la posibilidad de propiciar un aprendizaje incidental de conceptos estadísticos y asociarlos a nuevos conocimientos geográficos o reafirmar los que ya se tenían.

4.1 LOS DATOS

Antes de iniciar un análisis estadístico, se debe considerar de manera prioritaria la calidad de los datos básicos, así como el que estén completos y tengan homogeneidad. Sin la adecuada evaluación de los datos es seguro que ningún análisis por más complicado que sea pueda producir resultados que valgan la pena.

Se debe considerar entonces que los datos deben cumplir con dos procesos:

- a) Archivos brutos. Esto se refiere a la exportación de datos de Weatherlink al programa en el cual se realizará el análisis estadístico, en nuestro caso, hacia *Statística*.
- b) Validación y corrección. Ejecutar procedimientos que en la terminología meteorológica son conocidos como "control de calidad" (OMM, 1990).

En la realización de estos dos procesos así como en el resto del trabajo se utilizó el software estadístico *Statística/w* de Statsoft Inc. versión 4.5 (que también se emplea en la RAMA, el IMP y el CCA para el análisis de datos).

Al tener la necesidad de exportar datos de la estación hacia éste programa, se consideró lo siguiente:

- ❖ Weatherlink (el software de la estación) permite exportar sus datos mediante dos opciones:
 - Crear archivo texto
 - Crear archivo export
- ❖ *Statística* importa desde varios formatos como:
 - Excel
 - Lotus/Quattro
 - Dbase
 - Paradox
 - SPSS
 - ASCII
 - CSS
 - Megafiles Manager

Para detectar la opción más adecuada para el proceso de importación y exportación de datos y aún cuando esta actividad requirió dedicar mucho tiempo, se probó exportar datos de la estación en ambos formatos (texto y de exportación), el primer tipo mantiene el formato en que se generan reportes por impresora, sin embargo, no se recomienda usarlo para exportar los datos al programa estadístico ya que se requiere borrar y reformar el archivo en varios elementos (como la fecha). Mientras que con el formato de archivo de exportación, los datos se pueden importar fácilmente tanto por una hoja de cálculo como Excel (como se realizó inicialmente) o directamente por *Statística* (como se detectó que se podía hacer posteriormente) Figura 8.

Por cuestiones prácticas se grabaron 24 archivos de datos (correspondientes al período enero de 1997 a diciembre de 1998), uno por cada mes, sobre los que se realizaron posteriormente las pruebas de validación y análisis estadístico y que se detallan en las siguientes secciones.

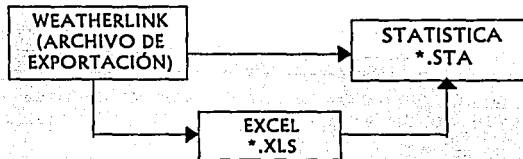


FIGURA 8. Dos posibilidades de exportar archivos de weatherlink

4.1.1 EXPORTACIÓN E IMPORTACIÓN DE DATOS

El formato nativo del software de la estación es *.BIN, por lo que, cuando se requiere utilizar los datos bajo otros programas es necesario transformarlos a un formato portable como es el caso de los archivos ASCII. Como se mencionó anteriormente el software meteorológico permite dos diferentes formatos para exportar sus datos; se describe enseguida el proceso de importación que se debe seguir en Excel con cada tipo de archivo, así como su contenido y las diferencias entre ellos.

- 1) Como ARCHIVO DE TEXTO la información se presenta con el siguiente formato:

ENP 3
01.01.97

TIEM	TAD	TAFU	MA	MIN	BAR	HAD	HAFU	ROC	VIEN	RACH	DIR	TERMI	LLUV
12:00am	16.4	11.6	11.8	11.3	788.9	43	66	5.2	0.4	1.3	W	11.6	0.0
12:15am	16.4	11.2	11.3	11.2	788.8	44	67	5.2	0.9	1.8	SW	11.2	0.0
12:30am	16.3	11.1	11.2	11.0	788.6	44	68	5.3	0.9	1.8	SW	11.1	0.0
12:45am	16.3	10.9	11.1	10.8	788.7	45	68	5.3	0.9	1.3	SW	10.9	0.0
1:00am	16.3	10.6	10.8	10.3	788.5	45	71	5.3	0.4	1.3	SW	10.6	0.0
1:15am	16.3	10.2	10.3	10.1	788.6	45	71	5.2	0.4	0.9	SW	10.2	0.0
1:30am	16.3	10.1	10.1	10.1	788.6	45	71	5.1	0.4	1.3	SW	10.1	0.0
1:45am	16.3	9.9	10.1	9.8	788.6	45	72	5.1	0.0	0.4	SW	9.9	0.0
2:00am	16.2	9.6	9.8	9.4	788.5	45	73	4.9	0.0	0.4	SW	9.6	0.0
2:15am	16.2	9.3	9.4	9.3	788.5	45	74	4.8	0.4	0.9	SW	9.3	0.0

Donde :

TIEM	=	Hora de observación	TAD	=	Temperatura Interior
TAFU	=	Temperatura exterior	MA	=	Temperatura máxima
MIN	=	Temperatura mínima	BAR	=	Presión atmosférica
HAD	=	Humedad interior	HAFU	=	Humedad exterior
ROC	=	Punto de rocío	VIEN	=	Velocidad del viento
RACH	=	Velocidad máxima del viento	DIR	=	Dirección del viento
TERMI	=	Factor de enfriamiento por viento	LLUV	=	Lluvia

Como se puede apreciar, de esta manera se tienen 14 variables, además del nombre de la estación y la fecha del día, éstos como encabezado; la variable del tiempo es de tipo alfanumérico y con valores que consideran un horario de 12 horas.

Cuando Excel intenta abrir este tipo de archivos hay que elegir

Paso 1.

Tipo de datos DE ANCHO FIJO

Asistente para importación de texto - paso 1 de 3

El asistente para conversión de texto estima que sus datos son de ancho fijo.
Si esto es correcto, elija Siguiente, o bien, el tipo que describa los datos con mayor precisión.

Tipo de los datos originales:

Elija el tipo de datos que los describa con mayor precisión:

Refinados Caracteres como comas o tabulaciones separan campos (estándar Excel 4.0)

De ancho fijo Los campos están alineados en columnas con espacios entre uno y otro.

Comenzar a importar en la fila: 1 Origen del archivo: Windows (ANSI)

Presentación preliminar del archivo C:\Mis documentos\0197.txt

2	ENP 3
3	01.01.97
4	
5	
6	TIEM TAD TAFU MA MIN BAR HAD HAFU ROC VIEN RAC
7	

Cancelar < Anterior Siguiente > Terminar

FIGURA 9. Paso 1 de 2 para la importación de archivos con formato texto

TESIS CON
FALLA DE ORIGEN

Paso 2.

Ampliar hacia la izquierda el campo de la variable ROC (punto de rocío), para evitar que los datos negativos pertenecientes a esta variable invadan la variable HAFU. Terminar.

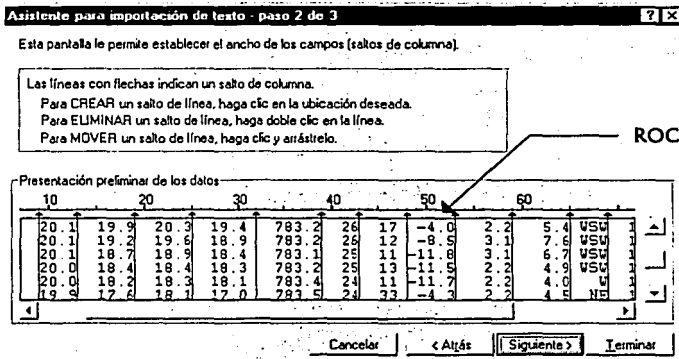


FIGURA 10. Paso 2 de 2 para la importación de archivos con formato texto

II) Como **ARCHIVO DE EXPORTACIÓN** los datos de la estación tienen el formato siguiente:

"Fecha", "Tiempo", "temperatura adentro", "temperatura afuera", "temperatura máxima afuera", "temperatura mínima afuera", "barómetro", "humedad adentro", "humedad afuera", "rocío", "velocidad del viento", "racha", "dirección del viento", "sensación térmica", "lluvia", "período"

"01.01.97", 0,16.4,11.6,11.8,11.3,788.9,43,66.5,2,0,4,1.3,12,11.6,0,0,15
 "01.01.97", 15,16.4,11.2,11.3,11.2,788.8,44,67.5,2,0.9,1.8,10,11.2,0,0,15
 "01.01.97", 30,16.3,11.1,11.2,11.0,788.6,44,68,5,3,0.9,1.8,10,11.1,0,0,15
 "01.01.97", 45,16.3,10.9,11.1,10.8,788.7,45,68,5,3,0.9,1.3,10,10.9,0,0,15
 "01.01.97", 100,16.3,10.6,10.8,10.3,788.5,45,71,5,3,0.4,1.3,10,10.6,0,0,15
 "01.01.97", 115,16.3,10.2,10.3,10.1,788.6,45,71,5,2,0.4,0.9,10,10.2,0,0,15
 "01.01.97", 130,16.3,10.1,10.1,10.1,788.6,45,71,5,1,0.4,1.3,10,10.1,0,0,15
 "01.01.97", 145,16.3,9.9,10.1,9.8,788.6,45,72,5,1,0.0,0.4,10,9.9,0,0,15
 "01.01.97", 200,16.2,9.6,9.8,9.4,788.5,45,73,4,9,0.0,0.4,10,9.6,0,0,15
 "01.01.97", 215,16.2,9.3,9.4,9.3,788.5,45,74,4,8,0.4,0.9,10,9.3,0,0,15

Se nota que en lugar de 14 variables son 16, pues a la fecha (día-mes-año) se le asigna el lugar de una variable y también se agrega la variable período, la cual indica en lapsos de cuántos minutos se realizan las mediciones. El formato de la variable "tiempo" tiene un formato que debe ser modificado. Otra diferencia es que la dirección del viento se maneja en formato numérico en lugar de ser alfabético, sin embargo, esto

facilita el manejo de esta variable con una doble asignación que puede manejar *Statistica/w*, para las variables nominales¹¹.

Ahora se siguen los siguientes pasos:

Paso 1.

Eligir tipo de datos DELIMITADOS

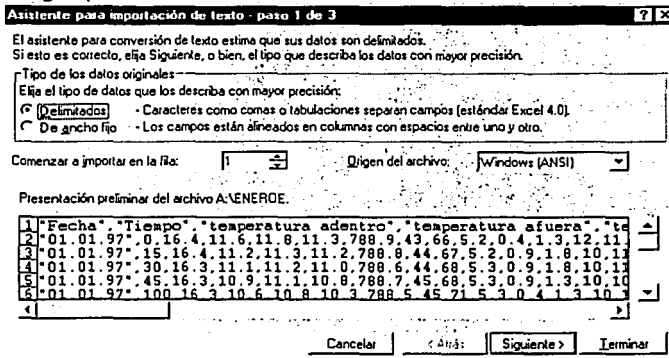


FIGURA 11. Paso 1 de 2 para importar archivos con formato de exportación Paso 2.

Aumentar como separador la coma e ignorar el siguiente paso. Terminar

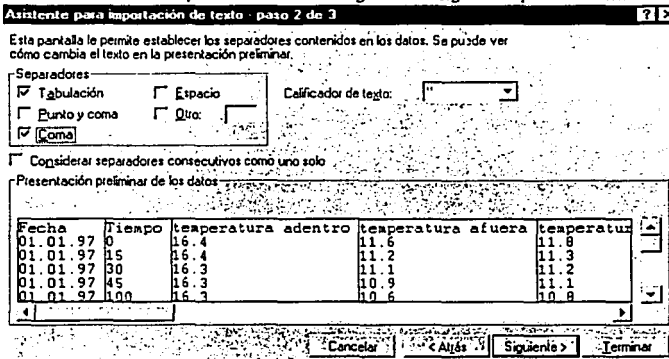


FIGURA 12. Paso 2 de 2 para importar archivos con formato de exportación

¹¹ Como en el manual no se especifica este cambio en la variable, se examinaron y compararon los archivos de texto y de exportación para deducir la relación entre el valor texto (16 rumbos de la rosa de los vientos) y su respectivo valor numérico.

Por último, se puede simplificar este proceso manejando directamente los archivos de exportación generados por Weatherlink y las opciones de Importación del programa *Statística*. Para facilitar este trabajo en la estación, se elaboró un macro que realiza este proceso y que se puede consultar en el *MANUAL*. Solo se requiere:

1) Que el usuario salve el archivo mensual que desea exportar, en la siguiente dirección y con el nombre:

C:\stat\Import\Import

2) Debe ingresar al módulo para manejo de datos y archivos del programa, eligiendo en el módulo Switcher de *Statística* la opción "Data Management/MFM" y el botón

Switch To

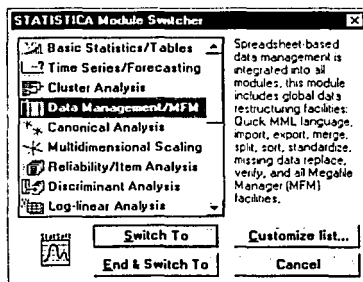


FIGURA 13. Módulo Switcher de *Statística*

3) Ejecutar el macro mediante la combinación de teclas **Ctrl + I**.

4) Asignar un nombre al archivo que se generó y salvarlo.

El archivo finalmente se presenta en *Statistica* de la siguiente forma:

Case	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11
1	01.04.97	15	22	15	16	15	789	40	64	9	1
2	01.04.97	0	22	16	16	16	789	40	64	9	1
3	01.04.97	15	22	15	16	15	789	40	64	9	1
4	01.04.97	30	22	15	16	15	789	40	65	9	0
5	01.04.97	45	22	15	15	15	789	40	65	9	0
6	01.04.97	100	22	15	15	15	789	40	66	9	1
7	01.04.97	115	22	15	15	15	789	40	65	9	0
8	01.04.97	130	22	15	15	15	789	40	63	9	0
9	01.04.97	145	22	16	16	15	789	40	69	9	0
10	01.04.97	200	22	15	15	15	789	41	69	10	0
11	01.04.97	215	22	15	15	15	789	41	65	9	0
12	01.04.97	230	22	15	15	15	789	41	69	9	0
13	01.04.97	245	22	15	15	14	789	41	73	10	1
14	01.04.97	300	22	14	14	14	789	41	73	9	0
15	01.04.97	315	22	14	14	14	788	41	72	9	1
16	01.04.97	330	22	14	14	14	788	42	73	9	0
17	01.04.97	345	22	14	14	14	788	42	72	9	0
18	01.04.97	400	22	14	14	14	788	42	72	9	0
19	01.04.97	415	22	14	14	14	788	42	69	9	0
20	01.04.97	430	22	14	14	14	788	42	67	9	0

FIGURA 14. Formato que presenta el archivo importado por *Statistica*

Al tener los datos en este formato, podemos manipularlos, validarlos y analizarlos de muy diversas maneras y de forma sencilla. Pero hay que adecuar el archivo para que se pueda trabajar con él, tanto numérica como gráficamente. Para esto se debe:

- ◆ Borrar el primer caso.
- ◆ Modificar la variable correspondiente a la hora (dividiendo entre 100) y eliminar la referente a "período" (última variable).
- ◆ Asignar a la variable de dirección del viento el correspondiente valor texto al valor numérico, tomando como referencia la Tabla III.

TABLA III. *Relación entre texto-número- grados para la dirección del viento*

DIRECCIÓN EN LA ROSA NAUTICA	VALOR EN ARCHIVO DE IMPORTACIÓN	EQUIVALENCIA EN GRADOS O ARCOS
N	0	360
NNE	1	22.5
NE	2	45
ENE	3	67.5
E	4	90
ESE	5	112.5
SE	6	135
SSE	7	157.5
S	8	180
SSW	9	202.5
SW	10	225
WSW	11	247.5
W	12	270
WNW	13	292.5
NW	14	315
NNW	15	337.5

- ◆ Redefinir los nombres de las variables, identificando previamente de qué tipo es cada una de ellas, partiendo de que en una serie climatológica las variables pueden ser:
 - a) **Discretas:** si el número de valores que puede tomar es contable (finito o infinito) y "si éstos se pueden arreglar en una secuencia que corresponde a los enteros positivos" Canavos (1988).
 - b) **Continuas:** "si sus valores consisten en uno o más intervalos de la recta de los reales" Canavos (1988).

Cabe hacer notar que en la realidad la dirección del viento es una variable continua pero se le convierte en discreta y categórica por la manera en que es registrada por los instrumentos de medición. En la siguiente tabla se ven los elementos climatológicos, los nombres asignados en los archivos y de qué tipo de variable se trata.

TABLA IV. Definición de las variables climatológicas y su tipo

	NOMBRE	FORMATO	TIPO
FECHA (juliana)	FECHA	MM-DD-AA	CONTINUA
HORA	HORA	NUMÉRICO	CONTINUA
TEMPERATURA INTERIOR	T INT	NUMÉRICO	CONTINUA
TEMPERATURA EXTERIOR	T EXT	NUMÉRICO	CONTINUA
TEMPERATURA MÁXIMA	T MAX	NUMÉRICO	CONTINUA
TEMPERATURA MÍNIMA	T MIN	NUMÉRICO	CONTINUA
PRESIÓN ATMOSFÉRICA	PRES ATM	NUMÉRICO	CONTINUA
HUMEDAD INTERIOR	HUM INT	NUMÉRICO	CONTINUA
HUMEDAD EXTERIOR	HUM EXT	NUMÉRICO	CONTINUA
PUNTO DE ROCÍO	P ROCIO	NUMÉRICO	CONTINUA
VELOCIDAD DEL VIENTO	VEL VTO	NUMÉRICO	CONTINUA
VELOCIDAD MAX. VIENTO	RACHA	NUMÉRICO	CONTINUA
DIRECCIÓN DEL VIENTO	DIR VTO	TEXTO/NÚMERO	DISCRETA
FACTOR ENFRIAMIENTO POR VIENTO	ENF_VTO	NUMÉRICO	CONTINUA
LLUVIA	LLUVIA	NUMÉRICO	CONTINUA

- ◆ Elegir la muestra horaria, es decir, elegir únicamente las observaciones de las horas 1...24
- ◆ Asignar nombre a los casos en donde se da inicio a cada día del mes, esto con el fin de facilitar el manejo gráfico.

Para realizar estos "arreglos" en el archivo de manera sencilla para el usuario de la estación se creó otro macro que se debe ejecutar inmediatamente después del macro realizado para el proceso de importación, el cual se corre dentro del mismo módulo de manejo de datos, mediante la combinación de teclas Ctrl + A y así el archivo tendrá las condiciones descritas previamente. El macro igualmente puede ser consultado en el anexo correspondiente al MANUAL. Por último el usuario solo deberá realizar manualmente el siguiente paso.

- ◆ Establecer correctamente la variable FECHA, de hecho se reemplaza pues al importar los archivos de la estación al programa *Statistica*, éste asigna automáticamente un valor numérico a cada valor texto que encuentra, por lo que, a la primera fecha de cada uno de los veinticuatro archivos ("01.01.97"... "01.12.98") les da el mismo valor (101) como se verifica al ver los valores texto que tiene asignado cada valor de fecha en la siguiente figura

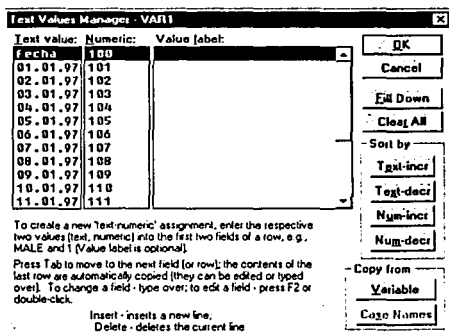


FIGURA 15. Asignación automática de valores numéricos a las fechas (texto)

Esto representa un problema, cuando se requiere reunir en un solo archivo la información de un año, e incluso de dos meses (sobre todo en procesos de ordenamiento cronológico). Para evitar este problema, el programa maneja un tipo de datos específico para el manejo de fechas, el cual es de tipo Juliano, o sea, que indica el número de días que han transcurrido en esta Era (d. de J.C.), y que a su vez maneja 6 formatos diferentes para su representación; otra ventaja de usar este tipo de fecha es que al cambio de siglo no hubo confusión.

Para convertir la fecha a formato Juliano, y así evitar cualquier problema posterior; se deben sustituir todas las fechas que originalmente tienen un formato de texto por otro que sea reconocido como uno de los 6 tipos fecha (DATE) del programa. Ejemplo: la fecha "01.01.97" que tiene asociado un valor numérico de 101, se cambia por 01/01/97. Se corrige la primera celda que contiene la fecha y se comprueba que valor se le asignó. Haciendo doble clic en el encabezado de la variable se despliega el cuadro que sigue

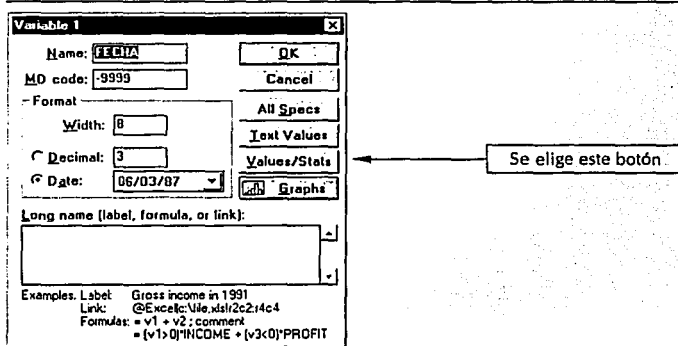


FIGURA 16. Diálogo para especificar una variable (CURRENT SPECS)

y verificamos en el siguiente cuadro los valores asignados automáticamente por *Statística*

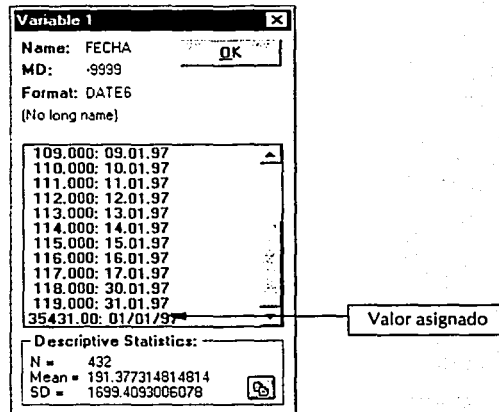


FIGURA 17. Valores que toma una variable

Así, utilizamos la función para recalcular variables, en este caso la variable FECHA



FIGURA 18. Menú de opciones para el manejo de variables. Recalcular variable

que a su vez utiliza el cuadro siguiente en donde se pueden definir 16 valores diferentes, aquí un ejemplo de cómo se cambiaron los primeros 16 valores (días del mes) asignados, por su correspondiente valor en tipo Juliano

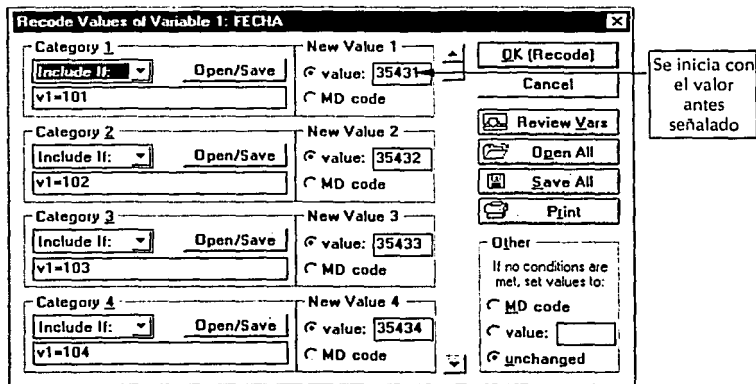


FIGURA 19. Diálogo para recalcular valores de una variable

realmente se realiza esta operación dos veces para cambiar todas las fechas de un archivo mensual (30 ó 31 días).

De esta forma se tiene el archivo listo para revisarlo, encontrar y eliminar posibles errores.

4.1.2 VALIDACIÓN

No se deben considerar aptos para el análisis, los datos que no han sido sometidos a un "control de calidad"¹², que asegure que los valores sean confiables y que exista continuidad; además esto permite detectar o descartar problemas tanto en los Instrumentos como en su operación. Este es un punto al cual se le han dedicado muchas horas/mujer, ya que existen varias pruebas de validación [ver OMM(1990)] a las cuales se ha sometido la base de datos, y que se detallan enseguida. Es importante recalcar que si esto no se hubiera realizado, las estadísticas obtenidas serían inútiles; por lo que se recomienda realizar esta actividad de manera regular y sistemática en la estación. Próximamente la estación será inspeccionada para verificar su funcionamiento (como recomienda la OMM que se haga cada 3 años) de manera que se tendrá disponible información referente a la calidad de los datos climatológicos. Sobre todo partiendo de una muestra uniforme respecto al horario de observación.

4.1.2.1. PRUEBAS DE IDENTIFICACIÓN E INTEGRIDAD

Esto se refiere al ordenamiento cronológico de los datos, así como la confirmación de que no existan datos alfabéticos en un campo definido como numérico y la identificación de datos perdidos o en blanco. A este respecto, podemos decir de acuerdo a lo que se logró observar en los archivos de la estación meteorológica, que sí están ordenados cronológicamente pero algunos meses no tienen la información completa. Después de realizar todas las pruebas de validación se elaboró una bitácora de observaciones¹³ para cada año (1997 a septiembre de 2001), que son mostradas en las Tablas VI a X respectivamente.

Podemos apreciar que para el año 1997 (consultar Tabla VI), existe un total de 466 observaciones que no fueron realizadas (o fueron eliminadas por ser erróneas, como se verá en las pruebas de tolerancia) lo que significa un 5.3% de datos "perdidos".

Por otra parte se uniformizó el horario de las muestras, pues en algunos meses de 1997 (enero y febrero) las observaciones se registraban cada 15 minutos, pero en el resto de ellos era cada 30 minutos y finalmente se estableció (por recomendación del Dr. Jauregui) que solo se debían considerar las muestras horarias; así que todas las estadísticas que se obtengan corresponderán a este último período de muestra. Para lograr lo anterior se automatizó esta tarea mediante la creación de un macro que agiliza la operación (ver macro "Adecuación al archivo importado" inciso 6, en MANUAL). Dichos macros pueden ser utilizados en archivos de otras estaciones que presenten la misma falta de uniformidad en los horarios de registro. Fue necesario reescribir para el

¹² En documentos de la OMM, se asigna este nombre al proceso de validación (ver OMM (1990) p. 4.23).

¹³ Verificar en 4.2 que una observación meteorológica es la medición de uno o varios elementos del clima, en adelante se usará este término para indicar que se tiene registro horario de todas o la mayoría las variables de la Tabla I.

período 30 de abril a 6 de mayo la variable hora ya que las observaciones se realizaron en horarios diferentes a como se venían realizando. Por ejemplo, en lugar de tomar la muestra a las 13:00 horas, se registraba 7 minutos antes – a las 12:53- en lapsos de 30 minutos. El 10 de septiembre también hubo un período en que además de que no hubo registro, se modificó el horario pues de las 5:30 siguieron las siguientes observaciones 10:42, 11:12... 16:12, 16:30, 17:00.

En lo que corresponde a 1988 (verificar Tabla VII), 135 observaciones no se realizaron (1.5%). En enero y hasta el 17 de febrero el período entre una observación y otra fue de 30 minutos, posteriormente se hizo cada hora, hasta el 27 de marzo en que se volvió a cambiar a lapsos de 30 minutos. Y finalmente desde mediados de abril hasta diciembre se estableció registrar las observaciones cada hora. Se reescribieron algunos datos pues el horario estaba desfasado en el mes de abril (los días 20 y 21) y en agosto (27 y 28), de hecho el archivo del día 27 indicaba los siguientes horarios 9:00, 10:59, 12:00, 11:00, 10:02, 11:02, 12:02, etc.

Durante 1999 solo se obtuvieron datos correspondientes al período de enero al 16 de abril, los demás datos se perdieron debido a la huelga. En el año 2000 se cuenta con datos a partir de mediados de febrero. Por último en el año 2001 se estudiaron los datos hasta el mes de septiembre, lo cual se puede verificar en las bitácoras de cada año (ver Tablas VIII, IX y X).

Existe también el problema de duplicidad de la última observación de un día y la primera del día siguiente, por lo que se eliminó la segunda de ellas.

4.1.2.2. PRUEBAS DE CONSECUENCIA INTERNA

Esta prueba se refiere a que los parámetros también pueden ser controlados por medio de comparaciones entre variables, por ejemplo se compararon de manera numérica, las variables referentes a temperatura las cuales deben cumplir la siguiente condición $T_{MIN} \leq T_{EXT} \leq TEM_{MAX}$ para considerarse correctas.

Existen variables como el punto de rocío, el factor de enfriamiento por el viento y la humedad relativa que son calculadas con base en otras variables y para las cuales existen formulas que sería conveniente verificar con el fabricante o distribuidor del equipo; pues no concuerdan los valores registrados con lo que se obtiene aplicando las fórmulas que menciona la literatura referente al tema (específicamente con la variable del factor de enfriamiento por viento).¹⁴

Otra comparación que se hizo, basándonos en la propia definición de los elementos fue $VEL_{VTO} \leq RACHA$.

Otra posibilidad es utilizar las gráficas de dispersión para verificar las relaciones que empíricamente¹⁵ se han comprobado, existen entre varias variables del clima.

¹⁴ Editor Qian Ping, China Agricultural Sciencetech Press. *Environment and Biometeorology*. P. 662

¹⁵ Ver sección "Análisis Multivariado" OMM (1990) p. 5.26

4.1.2.3. PRUEBAS DE TOLERANCIA

La mayoría de las variables que son registradas en la estación meteorológica tienen límites de registro, en primera por el propio Instrumento que los sensa (pues el fabricante informa de valores máximos y mínimos) y en segunda, porque hay límites naturales que determinan que un valor fuera de ellos sea improbable de ocurrir en la Ciudad de México, por ejemplo: la temperatura exterior, aún cuando el sensor puede detectar hasta los 60° C, no podría considerarse válido un valor alrededor de 40 °C aún en el verano.

De tal manera que conjuntando los límites Instrumentales y naturales, y utilizando las facilidades que proporciona el software estadístico para esta actividad se realizó esta validación. Fue necesario investigar los valores extremos que han ocurrido con anterioridad en el Distrito Federal, se tomaron como referencia las normales climatológicas y los valores extremos correspondientes a las estaciones meteorológicas de San Juan de Aragón y Tacubaya perteneciente al Servicio Meteorológico Nacional, así como datos proporcionados en el CCA, los cuales se resumen en la Tabla V.

Algunos datos tuvieron que ser eliminados de la muestra, pues sobrepasaban estos rangos y son considerados como "errores" (ver TABLAS XI Y XII), esta prueba se realizó tanto en los archivos con observaciones horarias como en los originales para saber la totalidad de errores que tenía la base de datos.

TABLA V. Intervalos de tolerancia

NOMBRE DE VARIABLE	INTERVALOS DEL INSTRUMENTO DAVIS	INTERVALOS PARA LA ZONA
T INT	0°C A 60°C	ND
T EXT, T MAX, T MIN	-45°C A 60°C	-5.5 A 34°C (A)
PRES ATM	880 A 1080 Mb	766 A 780 Mb (T)
HUM INT	10% A 90%	ND
HUM EXT	0% A 95%	Anual 46-71% (T)
P ROCIO	-73°C A 60°C	ND
VEL VTO, RACHA	Hasta 78.2 m/s	28.125 m/s (CCA)
DIR VTO	16 rumbos ó 0-360°	16 rumbos ó 0-360°
ENF VTO	-92°C A 60°C	ND
LLUVIA	0 A 999.8 mm (Díaria)	0 - 57 ¹⁶ mm (T)

(A) Registro de estación Aragón (1971-1980)
 (T) Registro de estación Tacubaya (1951-1980).
 {CCA} Dato proporcionado en el CCA
 ND Dato no disponible

¹⁶ Máxima precipitación en una hora (considerando un lapso de 25 años dentro del periodo 1951-1980).

Debido a que en 1998 se presentó el fenómeno meteorológico del Niño, se excedió el límite histórico en cuanto a temperatura por lo que en la validación no se respetó lo señalado en la tabla anterior, por ejemplo el 10 de mayo de 1998 se registró una temperatura máxima de 34.8 °C en la estación de la ENP 3 y aún cuando excede a la normal señalada se consideró una observación válida, pues la RAMA y el SMN(WWW,SMN) Informaban que ese día había sido el más caluroso del siglo (ENP3, 1998); por lo que este valor se tomó como el máximo permitido para validar esta variable. La presión atmosférica en la estación ha alcanzado valores de 803.5 Mb y como no existen datos de la estación de San Juan de Aragón respecto a esta variable, no se pueden descartar aún cuando difiera con los datos de la estación de Tacubaya.

4.1.2.4. PRUEBAS DE CONSECUENCIA TEMPORAL

Cuando la serie climatológica esta ordenada cronológicamente, o como se diría en términos estadísticos, que representa una serie de tiempo, este tipo de pruebas es de particular importancia, ya que nos *permite detectar indicios de desvío de una observación a otra, comparándolas con las anteriores y posteriores* (análisis univariado).

Esta actividad por sí misma implica conocer y verificar el comportamiento en la naturaleza de los elementos climatológicos, y en la estación se podría utilizar como un medio para reafirmar los conceptos teóricos adquiridos en la materia de Geografía General (4º año); mediante la comprobación directa con datos reales de la estación (sin correr el riesgo de modificaciones), lo cual propiciará un aprendizaje significativo. Al mismo tiempo que se induce a un aprendizaje extracurricular de conceptos estadísticos muy sencillos con el empleo de la semántica de *Statística*. Sobre todo haciendo hincapié en el análisis de tipo gráfico ya que el material resultante es atractivo para el estudiante y se entiende con mayor claridad; para lograr esto el potencial gráfico y la sencillez del programa es una excelente herramienta.

Veremos como se puede realizar este trabajo con respecto a diferentes elementos climáticos.

TEMPERATURAS

En los archivos climatológicos de la estación se consideran cuatro tipos de temperaturas, como ya se ha mencionado son las temperaturas interior, exterior y la máxima y mínima en el período, cabe hacer la aclaración de que estas dos últimas no son de utilidad en la labor cotidiana del climatólogo además que en nuestro caso de estudio pueden no ser las verdícas debido a la transformación que tuvieron los archivos al pasar de muestras con períodos de 15 o 30 minutos a períodos de una hora, lo que implica que los archivos contienen solo los valores de temperatura máxima y mínima de los 15 o 30 minutos precedentes a la hora de muestra, gráficamente la relación entre la temperatura interior y la exterior es la siguiente:

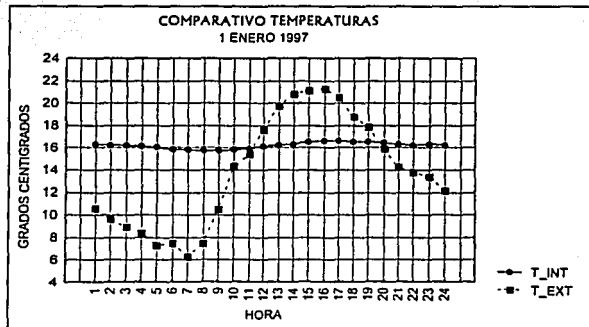


FIGURA 20. Comparación de temperaturas interior y exterior registradas en ENP 3

Además de ver la relación de estas variables podemos constatar lo que se menciona en la literatura meteorológica respecto al comportamiento de la variación en la temperatura a lo largo del día "la radiación nocturna produce un enfriamiento superficial de la tierra. El aire situado en la proximidad del suelo se enfría y se vuelve más pesado. Si el terreno está en pendiente, este aire frío desciende hacia niveles inferiores (viento catabático). Y durante el día se produce el fenómeno Inverso a causa del calentamiento de la pendiente (viento anabático); el aire más frío y por lo tanto más denso, viene a remplazar por abajo al que se eleva"¹⁷. También podemos apreciar que la temperatura en el interior de la estación, como es de esperarse, no presenta dichas variaciones.

Este comportamiento periódico también se puede apreciar en una gráfica mensual como se muestra aquí (24 observaciones x 31 días):

¹⁷ OMNI-SMN-SARH. *Compendio de apuntes para la formación del personal meteorológico. Clase IV. Meteorología.* México 1979.p 25 y 26

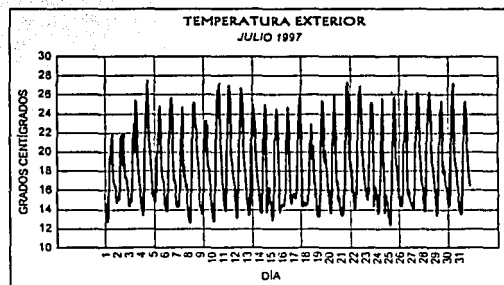


FIGURA 21. Comportamiento de la temperatura exterior en julio de 1997

Este tipo de gráfica mensual permite comparar cómo en un mes de primavera existe mayor variación en la temperatura que en otro de invierno. O se puede utilizar para confrontar su comportamiento con la información generada en el Servicio Meteorológico Nacional respecto a los frentes fríos que han transcurrido a lo largo de ese período.

Además, gracias al análisis gráfico, esta prueba nos permitió detectar un dato erróneo el día 13/06/97 en los datos de la estación, pues es evidente su desvío (ver TABLA XI).

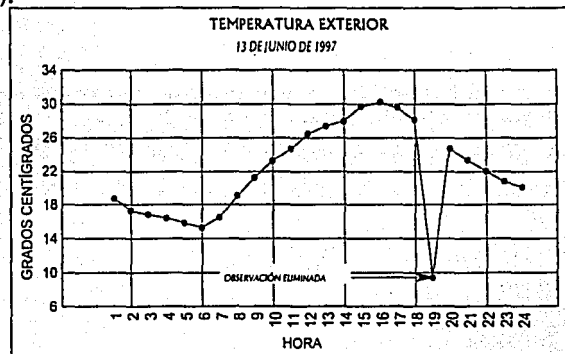


FIGURA 22. Representaciones gráfica ayudan a realizar pruebas temporales

PRESIÓN ATMOSFÉRICA

En un lugar determinado, este elemento varía continuamente y origina vientos y cambios en los demás elementos climatológicos. Estas variaciones pueden ser:

- Irregulares.- debidas principalmente al paso de sistemas de presión, así como su desarrollo o debilitamiento
- Regulares.- tienen períodos variados. Pero la oscilación más importante tiene uno de 12 horas aproximadamente por lo que se le conoce como "variación semidiurna".

Como lo vimos en el punto anterior, la sucesión de los días y las noches provoca alternativamente calentamiento y enfriamiento de la atmósfera, lo que produce oscilaciones rítmicas de expansión y contracción de la misma y que finalmente se traduce en variaciones de la presión. "Las presiones máximas se presentan a las 10:00 y 22:00 hr. locales aproximadamente mientras que las mínimas se producen a las 4:00 y 16:00 hr. locales aproximadamente. La variación semidiurna es un proceso bastante complejo"¹⁸ que podemos apreciar gráficamente de la manera siguiente:

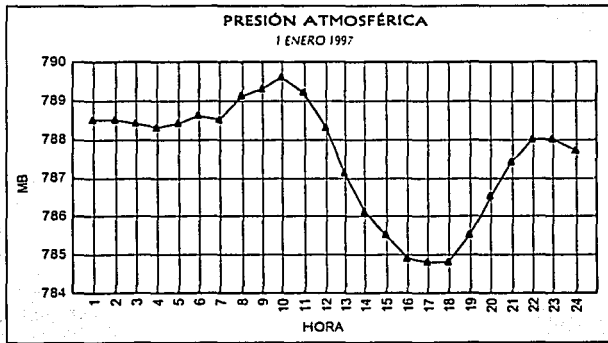


FIGURA 23. La gráfica comprueba la variación regular de la PRES_ATM

HUMEDAD RELATIVA

"Es la relación entre la masa de vapor de agua contenida en la unidad de volumen del aire y la del vapor de agua que sería necesario para saturar este volumen, a la misma temperatura"¹⁹. Esta expresada por la formula²⁰:

¹⁸ Op. Cit. OMNI-SMN-SARH. México 1979. P.38

¹⁹ Op. Cit. OMNI-SMN-SARH. México 1979. P.40

²⁰ Op. Cit. OMNI-SMN-SARH. México 1979. P.48

$$\text{Humedad relativa} = \left[\frac{\text{Tensión real del vapor}}{\text{Tensión del vapor saturante a la temperatura del aire}} \right] \times 100$$

Se expresa en tanto por ciento. Esta variable está relacionada a la temperatura ya que la tensión de vapor saturante disminuye cuando la temperatura decrece. Como se puede apreciar en la siguiente gráfica la humedad relativa generalmente alcanza sus valores máximos aproximadamente al amanecer, pues es entonces cuando se produce la temperatura mínima; durante el día esta última se eleva y provoca una disminución de la humedad relativa, lo que supone la desaparición de niebla o neblina.

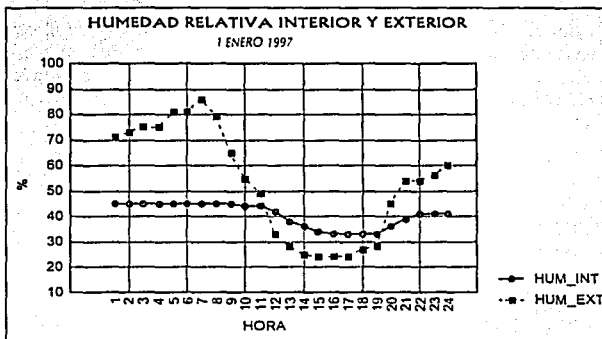


FIGURA 24. Gráfica que compara el comportamiento diario de humedades

VIENTO

Es el movimiento natural del aire atmosférico, generalmente tiene variaciones irregulares en período y magnitud, pues cambia rápida y continuamente.

“Aún siendo la naturaleza del terreno muy semejante, se observa a menudo una notable variación de la velocidad del viento en el transcurso del día. Alcanza su máximo entre el medio día y la caída de la tarde”²¹, pues desciende la temperatura, la convección disminuye y el viento también. En el alba alcanza su mínimo. Se considera que existe una racha cuando el viento aumenta bruscamente, respecto a su velocidad media, tomada en un cierto lapso de tiempo, dura menos de un minuto y va seguida de un debilitamiento.

²¹ Op. Cit. OMM-SMN-SARH. México 1979, P.57

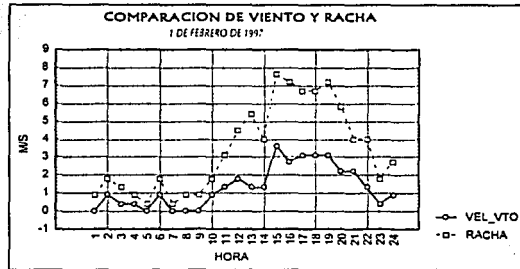


FIGURA 25. Gráfica que compara el comportamiento diario viento - rachas

Como se puede apreciar, los datos de la estación cumplen con lo que se cita en la teoría meteorológica y lo hemos confirmado con una simple representación gráfica, puede consultarse en el *MANUAL* la manera en que se crearon las Gráficas en dos dimensiones.

Gracias a la verificación de datos que contiene el programa estadístico algunas de las pruebas que explicamos previamente (especialmente las de tolerancia y las de consecuencia Interna) se realizan en un solo paso, mediante un macro desarrollado para este fin. Dichos límites de tolerancia (ver Tabla V), se guardaron en un archivo que puede ser recuperado en cualquier momento y que sirve para validar cada uno de los 24 archivos mensuales. El macro se ejecuta con la combinación de teclas **Ctrl+B** (ver *MANUAL*), y que se recomienda ejecutar después de los otros dos macros mencionados anteriormente, también dentro del módulo "Data Management".

Esta es la sintaxis para cada una de las condiciones establecidas en el archivo *LIMITES.TXT* (consultar en *MANUAL* macro de Validación de datos)²²:

Condición 1 (FECHA):	$v1 > = 35431$ and $v1 < = 36160$
Condición 2 (HORA):	$v2 > = 1$ and $v2 < = 24$
Condición 3 (T_INT):	$v3 > = -5.5$ and $v3 < = 34.8$
Condición 4 (T_EXT):	$v4 > = -5.5$ and $v4 < = 34.8$
Condición 5 (T_MAX):	$v5 > = -5.5$ and $v5 < = 34.8$
Condición 6 (T_MIN):	$v6 > = -5.5$ and $v6 < = 34.8$
Condición 7 (PRES_ATM):	$v7 > = 766$ and $v7 < = 804$
Condición 8 (HUM_INT):	$v8 > = 5$ and $v8 < = 95$
Condición 9 (HUM_EXT):	$v9 > = 0$ and $v9 < = 100$
Condición 10 (P_ROCÍO):	$v10 > = -20$ and $v10 < = 32$
Condición 11 (VEL_VTO):	$v11 > = 0$ and $v11 < = 28.2$

²² Para la presión atmosférica no se respetó lo señalado en la Tabla V, pues los valores que se presentaron en la estación fueron superiores a los de la zona de Tacubaya e inferiores a los instrumentales.

o ANÁLISIS ESTADÍSTICO

Condición 12 (RACHA):	$v12 > = 0$ and $v12 < = 28.2$
Condición 13 (DIR_VTO):	$v13 > = 0$ and $v13 < = 15$ and $v13 = -9999$ and $v13 = 255$
Condición 14 (ENF_VTO):	$v14 > = -20$ and $v14 < = v4$
Condición 15 (LLUVIA):	$v15 > = 0$ and $v15 < = 57$
Condición 16:	$v4 < = v5$ and $v4 > = v6$ and $v11 < = v12$ and $v8 < = v9$ (esta última es por la prueba de consecuencia interna)

A pesar de que se validó que la lluvia horaria no excediera lo señalado en la tabla V, al obtener el total anual de LLUVIA para 1997 se obtuvo un valor de 1074 mm y respecto a 1998 fue de 2278 mm, dichos valores sobrepasan de manera importante el valor de 749.6 mm que se consideró el año más lluvioso que se haya registrado en la estación Gran Canal durante el período (1972-1990); se comentó esto con la coordinación de la estación de la Preparatoria 3 y se verificó una falla física en el funcionamiento del pluviómetro el cual no vaclaba su contenido de manera correcta, por lo tanto se considera que los valores de LLUVIA no son totalmente confiables, sobre todo en los meses de verano cuando las lluvias son fuertes.

TABLA VI. Bitácora de observaciones durante 1997.

	MES											
	ENE	FEB	MZO	ABR	MYO	JUN	JUL	AGS	SEP	OCT	NOV	DIC
1												
2												
3												
4												
5												
6												
7							16					
8												
9					7							
10									6-16			
11												
12												
13						19		17				
14												
15										8		16-24
16				12-24								1-10
17	8-24											
18												
19												
20						7						
21				1-10								
22												
23						20						
24												
25												
26												
27												
28												
29												
30	1-7			10								
31												

Día con registro completo (cada hora)

Día sin registro


1-7 Horas sin registro


Días inexistentes en el calendario

TABLA VII. Bitácora de observaciones durante 1998

	MES											
	ENE	FEB	MZO	ABR	MYO	JUN	JUL	AGS	SEP	OCT	NOV	DIC
1												
2												
3												
4												
5												
6												
7												
8												
9												
10									17			
11												
12												
13												
14								8				
15				8								
16				7						18		
17									18			
18										17		
19												
20		8-24		8-10								
21												
22												
23												
24		1-8 11-24										
25		1-7	9, 11									
26												
27			8					10-12				
28												
29												
30												
31								20				

 Día con registro completo (cada hora)


 Día sin registro


 Horas sin registro


 Días inexistentes en el calendario

TABLA VIII. Bitácora de observaciones durante 1999

	MES											
	ENE	FEB	MZO	ABR	MYO	JUN	JUL	AGS	SEP	OCT	NOV	DIC
1												
2												
3												
4												
5				8								
6												
7												
8		14										
9			10									
10			19,20									
11												
12												
13												
14												
15												
16				17-24								
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
31												

 Día con registro completo (cada hora)

 Día sin registro

 1-7 Horas sin registro

 Días inexistentes en el calendario

ESTA TESIS NO SALI
DE LA BIBLIOTECA

TABLA IX. Bitácora de observaciones durante 2000

	MES											
	ENE	FEB	MZO	ABR	MYO	JUN	JUL	AGS	SEP	OCT	NOV	DIC
1	22-24		1-7									
2												
3												
4										7-24		
5												
6				7-24								
7												
8												
9												
10				1-6								
11												
12												
13												
14			8-24									
15		1-10	1-11									
16												
17												
18												
19												
20												
21									10			10-11
22												
23												
24								9, 17			1-11	
25												
26												
27												10
28											8-19	
29												
30											11-18	
31					9							

Día con registro completo (cada hora)

Día sin registro

1-7 Horas sin registro

Días inexistentes en el calendario

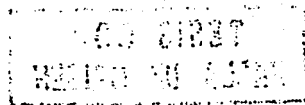


TABLA X. Bitácora de observaciones durante 2001

	MES								
	ENE	FEB	MZO	ABR	MYO	JUN	JUL	AGS	SEP
1									
2									
3									
4									
5									
6									
7									
8	11				8-24				
9					1-5				
10									
11						12-24			
12									
13								6	
14									
15						1-9			
16									
17									
18									
19						7-24			
20									
21								1-9	
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									

Día con registro completo (cada hora)

Día sin registro

1-7 Horas sin registro

Días inexistentes en el calendario

TESIS CON
FALLA DE ORIGEN

TABLA XI. Tabla de observaciones con algún error (Años 1997 Y 1998)

FECHA	HORA	T INT	T EXT	T MAX	T MIN	PRES ATM	HUM INT	HUM EXT	P ROCIO	VEL VTO	RACHA	DIR-VTO	ENF VTO	LLUVIA
13.01.97	8:45	16.1	8.3	8.7	8.0	789.1	19	30	-7.6	.4	40.2	ENE	8.3	0
30.01.97	13:15	22.2	39.4	-32768.0	20.1	788.8	28	27	16.9	1.3	4.0	NNW	39.4	0
30.01.97	13:30	22.5	251.5	-32768.0	20.2	788.6	27	26	-32768.0	.4	2.7	NNE	251.5	0
30.04.97	9:53	26.1	23.1	23.1	23.1	789.3	23	128	-32768.0	2.2	2.2	EN	22.5	92
13.06.97	19:00	30.8	9.4	27.1	26.5	786.9	19	19	-13.1	5.8	10.7	N	2.3	0
23.06.97	20:00	25.5	13.5	13.8	13.1	790.9	40	96	12.5	1.3	3.6	S	13.5	78
07.07.97	16:00	26.1	19.4	23.7	13.4	789.5	37	91	13.2	3.6	12.1	SE	17.2	99
13.08.97	17:00	25.4	15.5	17.9	14.3	790.6	38	77	10.7	3.1	8.0	N	13.2	67
10.09.97	10:42	22.2	20.9	20.9	20.9	790.8	46	52	10.7	.9	.9	SSW	20.9	225
25.03.98	9	21.2	17.3	18.9	15.4	800.6	19	22	-3.4	.4	114.0	N	17.3	0
25.03.98	11	22.9	22.2	24.2	20.9	800.4	21	19	-1.8	.9	114.0	N	22.2	0
31.08.98	20	22.4	15.9	16.1	15.8	794.2	61	96	14.8	2.2	5.4	EN	14.9	61
10.09.98	17	25.2	19.6	23.2	15.6	789.7	52	98	15.4	3.6	13.0	N	17.4	77
17.09.98	18	22.7	17.0	20.8	15.1	791.3	60	100	14.0	2.2	9.8	NNE	16.1	82
16.10.98	18	22.4	20.3	22.6	16.9	791.4	55	84	17.5	3.1	31.3	WSW	18.6	2
18.10.98	17	22.1	17.4	21.3	15.0	790.0	62	100	15.0	2.2	10.3	N	16.6	58

Los datos sombreados se consideran erróneos, por lo que se eliminó la observación (es decir todo el caso/renglón)

TABLA XII. Tabla de observaciones con algún error (Años 1999, 2000 Y 20001)

FECHA	OK	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT	TEXT
08.02.99	14	21.8	23.4	24.2	22.7	795.5	26	16	-3.4	3.1	114.0	N	22.2	0	
09.03.99	10	21.6	17.2	19.2	15.3	798.5	25	28	-3	1.3	114.0	NNE	17.2	0	
10.03.99	19	25.7	23.9	25.3	21.9	793.9	24	27	.9	17.9	48.3	E	18.7	0	
10.03.99	20	25.2	22.0	23.7	21.0	795.1	24	27	2.2	15.6	46.9	ESE	15.9	0	
10.03.99	22	24.9	19.4	19.9	19.0	796.2	27	29	.4	.9	43.4	ENE	19.4	0	
05.04.99	8	24.6	15.2	17.4	13.2	799.2	26	36	1.5	0.0	114.0	SSE	15.2	0	
31.05.00	9	21.9	18.0	20.8	16.9	801.2	53	128	32768	1.3	4.0	NE	18.0	0	
05.06.00	22	23.3	18.4	18.9	18.1	812.1	38	57	14.4	2.7	6.3	N	17.1	0	
05.06.00	24	23.0	16.9	17.1	16.7	811.3	39	100	16.9	1.8	4.9	NNW	16.9	0	
06.05.00	9	22.6	19.2	21.1	16.3	803.0	47	128	32768	2.7	4.9	N	17.9	0	
10.06.00	24	20.6	13.4	13.9	13.2	800.7	61	100	13.4	1.3	3.6	W	13.4	63	
12.06.00	4	20.1	14.6	14.8	14.4	799.9	64	128	32768	.4	1.3	W	14.6	0	
12.06.00	5	20.1	14.8	14.8	14.7	800.2	64	128	32768	0.0	1.3	W	14.8	0	
12.06.00	6	20.1	14.9	15.1	14.8	800.2	64	128	32768	.9	3.1	W	14.9	0	
17.06.00	9	20.9	19.8	21.6	17.6	800.0	60	128	32768	0.0	1.3	SW	19.8	0	
18.06.00	1	20.6	14.4	14.5	14.3	800.8	64	128	32768	1.3	3.1	SE	14.4	0	
28.06.00	11	22.0	24.1	25.4	22.8	801.9	44	128	32768	.9	3.6	NNE	24.1	1	
28.06.00	18	22.0	14.1	14.6	12.8	799.0	53	128	32768	2.2	6.3	NW	13.1	5	
28.06.00	19	21.6	15.1	15.7	13.9	799.8	55	128	32768	1.8	4.9	NNW	15.1	21	
28.06.00	20	20.9	13.2	14.0	12.7	800.7	58	100	13.2	3.1	8.9	N	10.7	77	
29.06.00	8	20.9	14.2	15.2	13.4	802.3	60	128	32768	1.3	3.1	SW	14.2	0	
21.09.00	10	21.6	20.7	22.3	19.9	796.4	57	128	32768	.9	2.7	ENE	20.7	0	
24.09.00	9	21.7	19.2	21.2	18.2	799.4	55	128	32768	.4	2.2	NNE	19.2	0	
24.09.00	17	22.7	24.7	25.3	24.1	795.3	46	128	32768	2.7	5.4	N	23.9	0	
08.05.01	5	19.3	10.1	10.0	10.0	805.7	45	94	8.3	4.5	13.9	N	4.3	5	
14.06.01	19	22.9	14.4	15.4	13.8	800.0	51	92	12.8	5.4	10.3	N	9.2	59	

Los datos sombreados se consideran erróneos, por lo que se eliminó la observación (es decir todo el caso/renglón)

4.2 ESTADÍSTICA DESCRIPTIVA

En esta sección se introducirán algunos conceptos básicos de estadística, aun así se recomienda consultar y apoyarse en algún texto de estadística elemental como los recomendados en el CCH Y ENP, para los cursos de la materia optativa Estadística y Probabilidad, a fin de tener más detalles de dichos conceptos.

Se ha intentado seguir los contenidos curriculares de la materia a nivel bachillerato para estar acorde con los conocimientos previos del estudiante o el coordinador, y se utiliza a *Statística* como herramienta de apoyo en la realización de diversas tareas de análisis de datos, tomando para los ejemplos dos archivos de la estación (el mes de Julio de 1997 - h0797.sta- y otro en que se incluyeron los 12 archivos mensuales 1997.sta), enfocando el interés en las variables de mayor importancia: T_EXT, T_MAX, T_MIN, PRES_ATM, HUM_EXT, VEL_VTO, RACHA, DIR_VTO y LLUVIA (se incluye por su importancia climatológica y solo para mostrar un ejemplo, pero no debe olvidarse que los datos no son confiables). Descartando las variables correspondientes al interior de la estación, ya que son de un uso limitado y tampoco se analizarán las variables en las que existe duda sobre su correcta medición (P_ROCIO, ENF_VTO).

En nuestra vida cotidiana y muchas veces sin darnos cuenta utilizamos uno de los principios fundamentales de la estadística, como ejemplo muy burdo: hemos escuchado frases como "Todas las mujeres son iguales", obviamente quien lo dice no conoce a TODAS las mujeres pero basándose en ALGUNAS experiencias hace una generalización. A este proceso de tomar como base algo específico para realizar una afirmación más general se le conoce estadísticamente como INFERENCIA. Claro está que por ser la estadística una ciencia, sus afirmaciones tienen fundamento en el método científico y son susceptibles de comprobarse.

Si seguimos estudiando estadísticamente el ejemplo anterior, al grupo de TODAS las mujeres se le nombraría POBLACIÓN y al subgrupo o subconjunto de ALGUNAS mujeres se le llamaría MUESTRA. Así es que una población incluye a la totalidad de los elementos que nos interesan, mientras que la muestra es solo una parte de aquella.

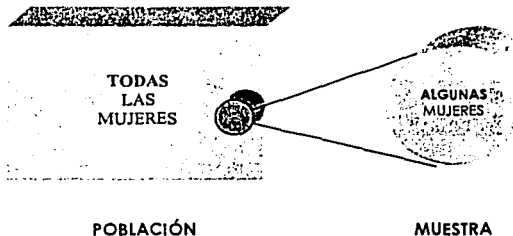


FIGURA 26. Relación entre población y muestra

La estadística pone en nuestras manos, mediante su metodología (ver 2.2), la posibilidad de estudiar una muestra y conocer de manera aproximada las características de la población de donde se extrajo. Esta es la principal finalidad de esta ciencia.

En la vida real, existen muchos ejemplos como el anterior en el que es prácticamente imposible conocer la población completa, como lo es el caso que nos ocupa.

Nos interesa conocer más detalles del clima que prevalece en la zona geográfica en que se ubica la preparatoria No.3, pero sabemos que esta conformado por muchos elementos climatológicos (ver 3.3) (estadísticamente constituyen las UNIDADES ELEMENTALES). Con ayuda del equipo Davis se pueden realizar y recopilar observaciones de ciertos elementos como la temperatura, la humedad, etc.; sin embargo, debido a su propia naturaleza no podemos tener un registro o conteo total de dichas observaciones, de modo que se dispone de una muestra formada por las observaciones horarias provenientes de una población infinita que es el clima real en el noreste del Valle de México.

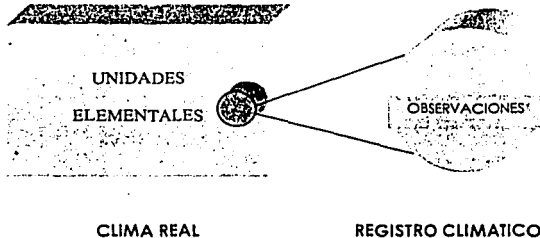


FIGURA 27. Relación entre población y muestra en el caso del clima

La estadística nos indica que para que una muestra sea útil debe ser:

- a) aleatoria, es decir, que las observaciones que la conformen sean elegidas al azar, asegurando así que todas tengan la misma posibilidad (probabilidad) de ser incluidas
- b) representativa, o sea, que refleje las características de la población de donde son originarias.

En la meteorología, la ingeniería y en general en las ciencias físicas, la población de interés no consiste de objetos tangibles a partir de los cuales se selecciona un cierto número para formar la muestra; sino que la población está formada por un número infinito de posibles resultados para alguna característica medible que nos interese, como puede ser la presión atmosférica, la humedad relativa, etc.

Se considera que el siguiente procedimiento es una forma de realizar un muestreo aleatorio (con reemplazo) para una población de este tipo: "Se diseña un experimento y se lleva a cabo para proporcionar la observación X_i de la característica medible X . El experimento se repite bajo las mismas condiciones proporcionando el valor X_i . El proceso

se continua hasta tener n observaciones X_1, X_2, \dots, X_n de la característica X " Canavos (1988).

Cuando se tiene una muestra de datos climatológicos, estos deben ser ordenados o clasificados de alguna forma legible y que pueden transmitir su significado de forma sencilla, ya sea en forma de cuadros o gráficas, que nos permitan apreciar sus características. De esta manera se calculan las medidas descriptivas como son los "promedios" (medias), proporciones o desviaciones estándar, a las cuales se les llama **ESTADÍSTICAS**. A la parte de la ciencia estadística que se dedica al cálculo de estas medidas se le llama **ESTADÍSTICA DESCRIPTIVA**, ella nos aporta las bases para aproximarnos (hacer estimaciones) de las características de la población llamadas **PARÁMETROS** o realizar prueba de hipótesis respecto a dicha población (**ESTADÍSTICA INFERENCIAL**).

El proceso simplificado se muestra en la siguiente figura, se ve que las estadísticas son valores numéricos reales que se calculan a partir de los datos de una muestra, y por medio de la inferencia tratamos de estimar los parámetros de la población, comúnmente "tratamos con uno o más parámetros que nos describen la población. No sabemos los valores de los parámetros (y por lo general nunca los sabremos)" Hunsberger (1983).

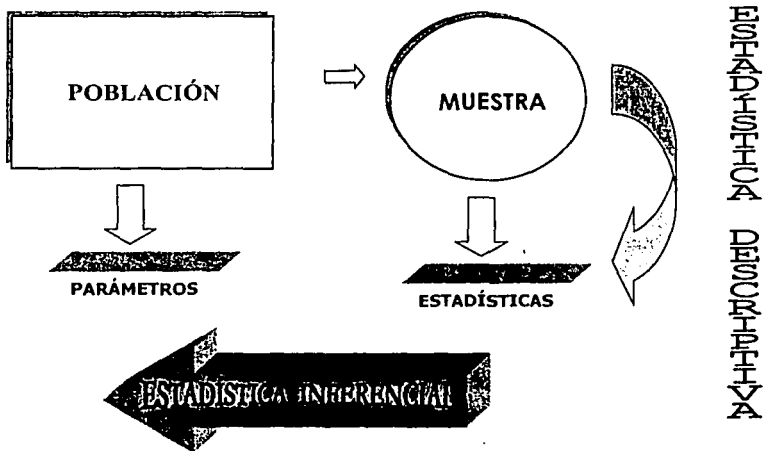


FIGURA 28. Metodología estadística

De esta manera, la estadística inferencial le permite al climatólogo hacer afirmaciones u obtener conclusiones respecto al clima de un lugar pero "es necesario usar

la probabilidad por que una conclusión basada en una muestra proviene de una información incompleta respecto a la población, y por lo tanto no puede hacerse con total certeza" Chou (1972). Esta es la razón de que la probabilidad y la estadística estén siempre involucradas, por ejemplo una afirmación estadística nos dirá así: La probabilidad de lluvia el día 30 de agosto de 1997 es de 0.95, con un error no mayor de 3%.

4.2.1 TABLAS DE FRECUENCIA Y SU REPRESENTACION GRÁFICA

Como se mencionó, motivados por el deseo de conocer más sobre el clima en la zona, se reúnen, clasifican y describen los datos climatológicos para identificar los patrones en nuestro conjunto de datos. La utilidad de esta clasificación llamada DISTRIBUCIÓN DE FRECUENCIAS, es que con un vistazo podemos apreciar de manera aproximada valores como el máximo y mínimo, el rango entre estos valores, en qué puntos se concentra el conjunto de datos, etc. Para organizar y analizar datos del clima, es muy frecuente el uso de las herramientas estadísticas que se mencionan en esta sección, debido a las limitaciones de espacio solo se presentarán algunos ejemplos de la manera en que se pueden aplicar al caso de las estaciones y con algunas variables específicas, sin embargo, para un estudio exhaustivo pueden consultarse las referencias meteorológicas y estadísticas mencionadas a lo largo del trabajo

Primero se organizan las observaciones muestrales en pequeñas CLASES²³ o intervalos de clase, estas deben ser mutuamente exclusivas, es decir, que no se traslapan entre sí, y colectivamente exhaustivas, esto es, que incluyan a todos los datos. Al número de observaciones incluido en una clase se le llama FRECUENCIA DE CLASE y se le denota por (*f_i*). A su vez una frecuencia de clase representa un cierto "porcentaje" del total de observaciones (*n*), a dicho porcentaje se le conoce como FRECUENCIA RELATIVA y su fórmula de cálculo es *f_i/n*, en un histograma de frecuencias relativas su área será siempre igual a 1.

La representación tabular de los datos es más compacta que la que tienen los datos brutos y pretende condensarlos o simplificarlos sin perder muchos detalles, se le da el nombre de TABLAS ó CUADROS DE FRECUENCIA, y sus equivalentes gráficos son los HISTOGRAMAS y los POLÍGONOS DE FRECUENCIA en estos se utiliza el eje cartesiano (XY), en el eje horizontal (X) se representan las clases en que son clasificados los datos mientras que en el eje vertical (Y) se muestra el número de observaciones o frecuencias (absolutas o relativas) para cada una de las clases.

Es importante elegir un número adecuado de clases, veamos porqué es esto, analizando el caso de la T_EXT en Julio de 1997 y como se realiza esto con *Statística*. Notemos cómo aunque estemos utilizando el mismo archivo de datos de temperatura se pueden representar de diferentes formas, conforme al número de clases elegidas.

²³ El número de clases debe ser como mínimo 5 y un máximo de 15. Existe una regla que indica que el número de ellas no debe exceder cinco veces el logaritmo decimal del número de observaciones. OMM(1990).

Utilizando el módulo *BASIC STATISTICS/TABLES* tenemos las siguientes opciones de operación, entre las que debemos elegir la correspondiente a las tablas o cuadros de frecuencia:

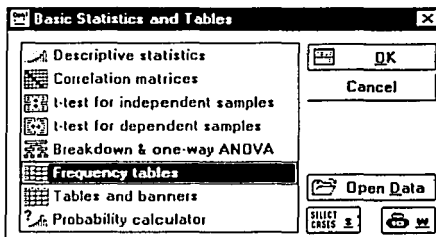


FIGURA 29. Cuadro inicial para el manejo de la estadística básica y tablas

Eliendo previamente la variable o variables a trabajar, ya sea, mediante la elección directa en la hoja de datos o posteriormente utilizando el diálogo de selección de variable (ver. M 13) que se presenta al elegir el botón **Variables:** del siguiente cuadro.

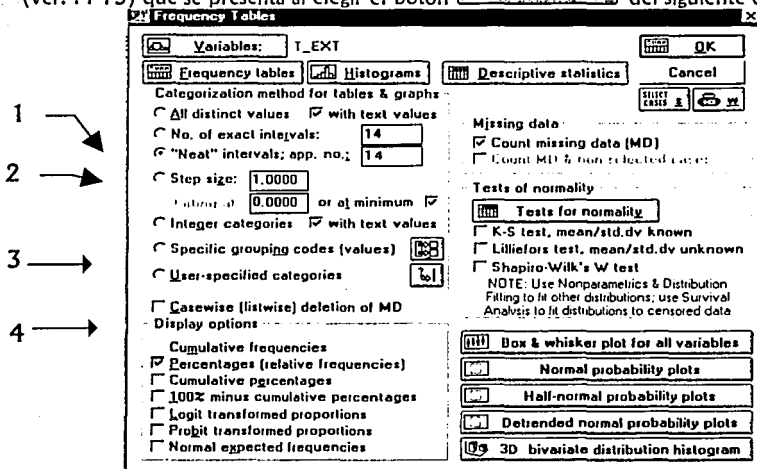



FIGURA 30. Permite la creación de tablas de frecuencia e histogramas

Como podemos ver aquí existe una sección "Categorization method for tables & graphs" que presenta siete opciones, de las que solo una se puede seleccionar y que determina la manera en que se crearán las clases, sobre las que se distribuirán los datos de la variable a analizar.


Mediante la opción marcada con el número 1 "No. of exact intervals" (seleccionamos 14 clases²⁴) podemos producir la siguiente tabla de frecuencias para la variable T_EXT, eligiendo el botón  Frequency tables

Category	Count	Percent
11.8231<=x<12.9769	11	1.48048
12.9769<=x<14.1307	65	8.74832
14.1308<=x<15.2846	134	18.03499
15.2846<=x<16.4384	109	14.67026
16.4385<=x<17.5923	71	9.55585
17.5923<=x<18.7461	62	8.34455
18.7462<=x<19.9000	51	6.86406
19.9000<=x<21.0538	38	5.11440
21.0538<=x<22.2076	50	6.72948
22.2077<=x<23.3615	47	6.32571
23.3615<=x<24.5153	35	4.71063
24.5154<=x<25.6692	46	6.19112
25.6692<=x<26.8230	19	2.55720
26.8231<=x<27.9769	5	.67295
Missing	0	0.00000

FIGURA 31. Tabla de frecuencia creada por la opción 1

En cuanto a la interpretación de la tabla de frecuencia :

- La primera columna nombrada como "CATEGORY" nos indica el intervalo de valores para cada clase, conocido como intervalo de clase.
- La segunda columna "COUNT" se refiere a la frecuencia de clase (f_i), por ejemplo, tomando como base la fig. 31, podemos afirmar que la mayoría de las observaciones de T_EXT en ese mes, se encontraron entre los 14.13 y 15.28 grados centígrados, 134 observaciones, para ser exactos, es decir, un 18.03% del total de la muestra (esto lo indica la tercera columna "PERCENT"²⁵).
- También podemos apreciar que el rango que tuvieron las temperaturas para julio de 1997 se encontró entre los 11.8°C y los 27.98°C aproximadamente.

Este cuadro tiene asociada como representación gráfica, el siguiente histograma de frecuencias que se obtiene al dar un clic en el botón  Histograms

²⁴ Siguiendo la regla de que el número de clases no debe ser mayor que cinco veces el logaritmo del número de observaciones, en este caso se tienen 743.

²⁵ Es el valor de la FRECUENCIA RELATIVA (f/n), multiplicada por 100.

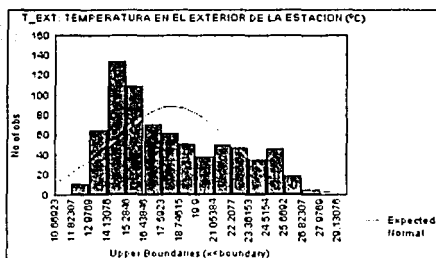


FIGURA 32. Histograma de frecuencias absolutas creado por la opción 1

veamos que ocurre con la opción marcada con el número 2 (los intervalos tienen valores con redondeo, para facilitar su lectura; sin embargo, no siempre respeta el número de intervalos solicitado, como se verá después de este cuadro y gráfica).

Category	Count	Percent
12.0000<=x<14.0000	65	8.74832
14.0000<=x<16.0000	218	29.34051
16.0000<=x<18.0000	125	16.82369
18.0000<=x<20.0000	98	13.18977
20.0000<=x<22.0000	75	10.09421
22.0000<=x<24.0000	74	9.95962
24.0000<=x<26.0000	70	9.42127
26.0000<=x<28.0000	18	2.42261
Missing	0	0.00000

FIGURA 33. Tabla de frecuencias creada por la opción 2 y su correspondiente histograma de frecuencias es el que sigue

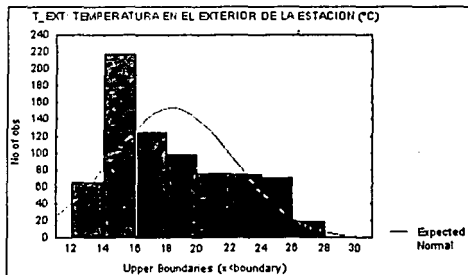


FIGURA 34. Histograma de frecuencias absolutas creado por la opción 2

Con estas 8 clases podemos saber que las temperaturas se concentran entre los 14 y 16 °C y no como en la anterior en que conocimos con mayor exactitud que la mayoría estaba entre los 14 y 15 °C.

Pero si utilizamos la opción 2 con un número de intervalos menor, por ejemplo el 8. Se genera la siguiente tabla de frecuencias (como se ve solo tiene 7 intervalos)

Category	Count	Percent
10.0000<=x<12.5000	1	.13459
12.5000<=x<15.0000	182	24.49529
15.0000<=x<17.5000	202	27.18708
17.5000<=x<20.0000	121	16.28533
20.0000<=x<22.5000	92	12.38223
22.5000<=x<25.0000	90	12.11306
25.0000<=x<27.5000	55	7.40242
Missing	0	0.00000

FIGURA 35. Tabla de frecuencias creada por la opción 2 con menos intervalos

y este es el histograma correspondiente

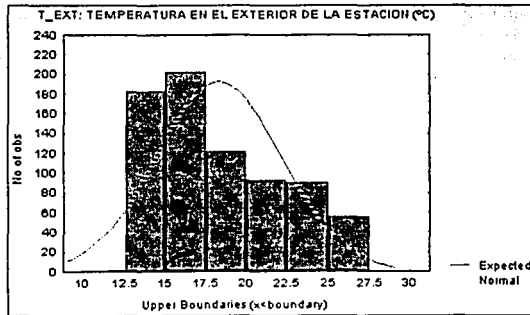


FIGURA 36. Histograma de frecuencias absolutas. Opción 2 con 8 intervalos

De modo que el programa estadístico nos permite probar diferentes formas para elaborar un histograma pero debemos decidir el tipo de gráfica más adecuado para lo que deseamos estudiar. Otra de las formas de representación de una distribución de

frecuencias, es un polígono de frecuencias, se trazan simplemente los puntos medios de cada clase y se unen por una línea recta. Como se ve a continuación:

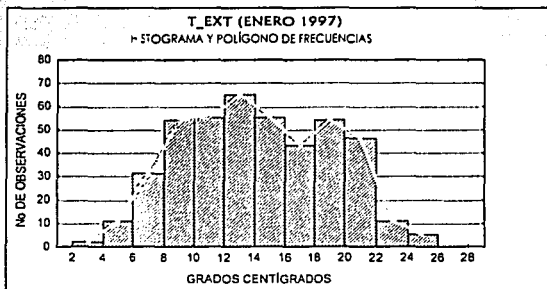


FIGURA 37. Histograma y polígono de frecuencias absolutas para la T_EXT.

El uso de los polígonos de frecuencias tiene dos ventajas, una que es más fácil comparar varias distribuciones en la misma gráfica, pues es más claro superponer polígonos que histogramas de frecuencia.

La otra ventaja es que los polígonos de frecuencia sugieren una curva suave como representación idealizada de la distribución de la población. Sabemos que la muestra solo tiene algunos elementos de la población y que por eso al representarlos gráficamente presentan irregularidades, pero si la muestra pudiera ser mayor y el número de intervalos menor se espera que la curva sea cada vez más suave y regular; ésta adquiere importancia pues se considera que representa la verdadera distribución de la población.

Los meteorólogos y matemáticos han detectado basándose en largas series de datos, las curvas a las que se aproximan varios elementos climáticos y pueden asumir muchas formas, podemos ver que los histogramas de las figuras anteriores tienen superpuesta -por *Statística*- una curva en forma de campana, la cual se llama distribución de probabilidad normal o Gaussiana, y que es la más usada y conocida en la labor estadística.

Ahora veamos los comportamientos de las curvas que generan los polígonos de frecuencias de otras variables, por ejemplo la presión atmosférica (las figuras 38 a 42 se generaron a partir del archivo anual de 1997, en el *MANUAL* se puede consultar la manera en que fue creado dicho archivo, así como la forma de elaborar este tipo de gráficas utilizando *Statística*).

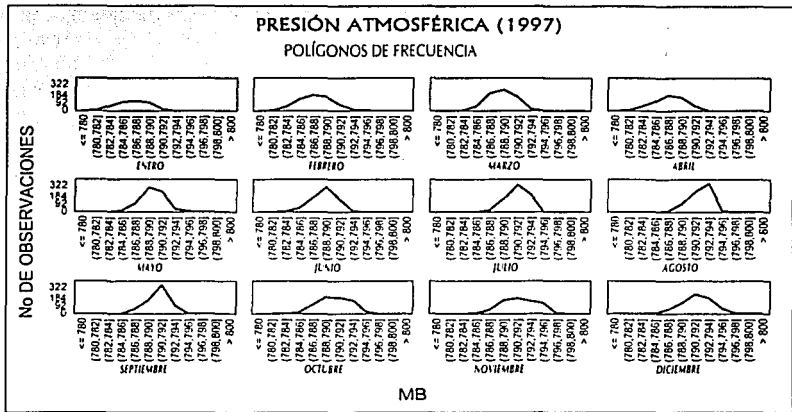


FIGURA 38. Polígonos de frecuencias para la PRES_ATM mensual de 1997

Podemos apreciar que se parece al tipo de curva normal, es decir, que se comporta casi simétrica y en forma de campana.

Veamos ahora que sucede con la variable HUM_EXT, durante 1997:

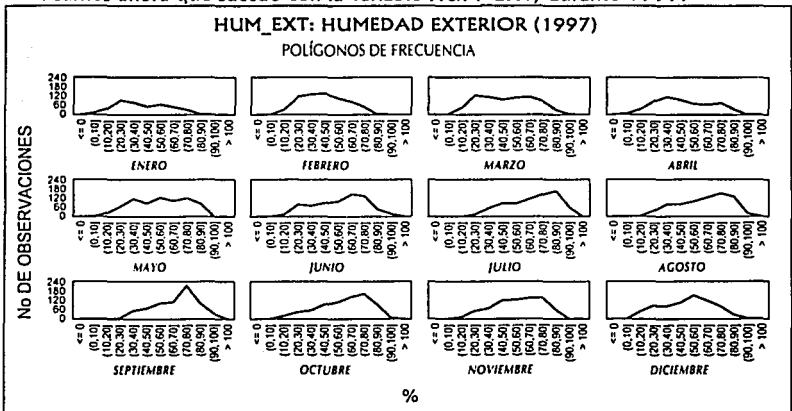


FIGURA 39. Polígonos de frecuencias absolutas para la HUM_EXT.

Como podemos apreciar en la figura la humedad relativa en el exterior tiene una pequeña desviación en casi todos los meses, se dice que esta forma de curva es **NEGATIVAMENTE ASIMÉTRICA** (la cola más larga de la curva esta del lado izquierdo).

Respecto a la velocidad del viento, veremos en la siguiente figura que presenta una distribución muy diferente:

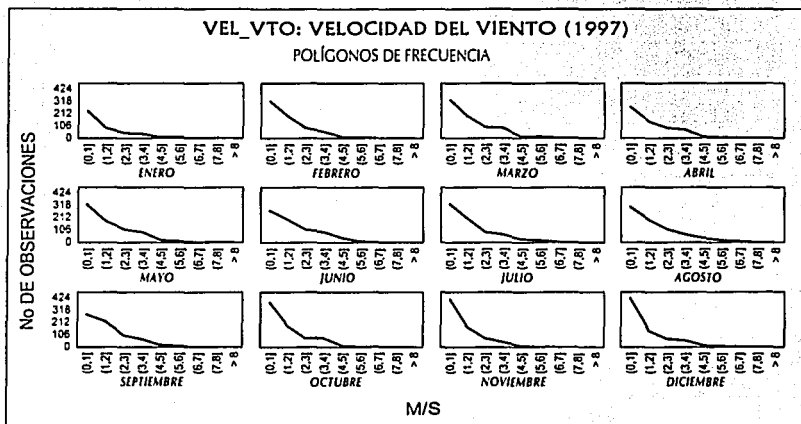


FIGURA 40. Polígonos de frecuencias absolutas para la VEL_VTO.

A esta forma de curva se le conoce como J invertida. Las velocidades que tienen los vientos, son en su mayoría de pocos m/s y solo algunos vientos son fuertes. Y existe cierto cambio en cuanto al comportamiento de las rachas de viento como veremos en la siguiente figura

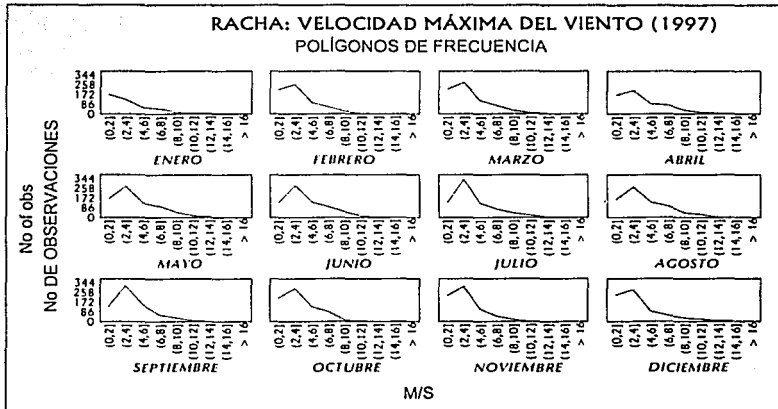


FIGURA 41. Polígono de frecuencias absolutas para la variable RACHA

Como la dirección del viento es una variable categórica ó nominal, no se representa mediante polígonos de frecuencia, sino que se utiliza el histograma de frecuencias para determinar la dirección dominante del viento, veamos por ejemplo lo que ocurrió para cada mes de 1997 con la siguiente figura.

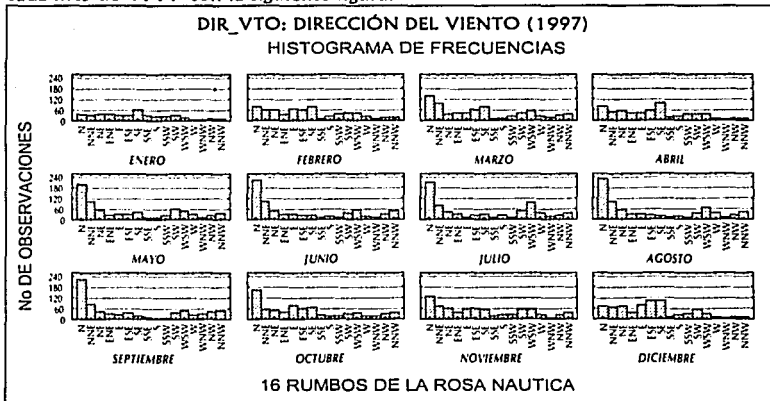


FIGURA 42. Histograma de frecuencias para la DIR_VTO.

**TESIS CON
FALLA DE ORIGEN**

Con esta gráfica podemos verificar mensualmente la dirección del viento y comprobar que son los vientos del norte los predominantes.

Como se mencionó, la dirección del viento maneja valores de tipo texto por lo que es recomendable utilizar la opción "ALL DISTINCT VALUES" ó la opción señalada con el número 3 "INTEGER CATEGORIES" (ya que a cada texto se le asignó un valor del 1 al 16), para la construcción de la tabla de frecuencias e histograma correspondiente.

En cuanto a la variable de LLUVIA, podemos partir de construir una tabla de frecuencias conforme la clasificación de lluvias que maneja el SMN (en milímetros):

- Escasas < 5 mm
- Ligeras ≥ 5 y < 10 mm
- Moderadas ≥ 10 y < 20 mm
- Fuertes ≥ 20 y < 50 mm
- Muy fuertes ≥ 50 y < 70 mm
- Intensas ≥ 70 mm

Para esto hay que utilizar la opción marcada con el número 4 "USER_SPECIFIED CATEGORIES", al elegir esa opción aparece un cuadro de diálogo para la definición de las categorías o clases, como se ve en el siguiente cuadro en que ya se han seleccionado las clases ha crear (no se visualiza la última categoría pero se definió como "LLUVIA >= 50 AND LLUVIA < 70").

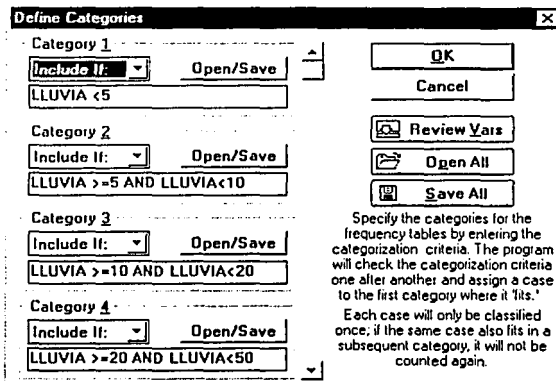


FIGURA 43. Cuadro para definir los tipos de lluvia

Y obtendremos la siguiente tabla de frecuencias:

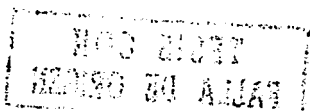


TABLA XIII. *Tabla de frecuencias para LLUVIA en julio de 1997*

	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA	FRECUENCIA RELATIVA ACUMULADA
LLUVIA < 5	727	727	97.85	97.85
LLUVIA ≥ 5 AND LLUVIA < 10	7	734	.94	98.79
LLUVIA ≥ 10 AND LLUVIA < 20	4	738	.54	99.33
LLUVIA ≥ 20 AND LLUVIA < 50	4	742	.54	99.87
LLUVIA ≥ 50 AND LLUVIA < 70	1	743	.13	100

No existe la opción para generar el histograma pero esta sería su representación:

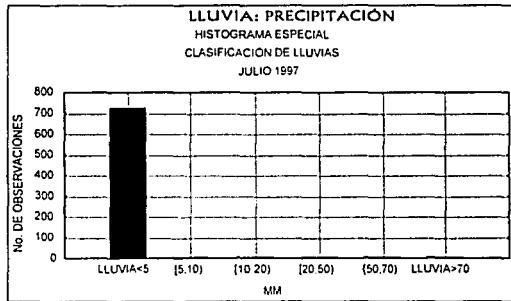


FIGURA 44. *Histograma de frecuencias para la LLUVIA en julio de 1997*

Similar al anterior histograma es el de los 11 meses restantes, pues las precipitaciones que se dan en el Distrito Federal son generalmente escasas y en los meses de verano solo hay unas cuantas lluvias fuertes.

Como podemos ver, el uso de las tablas de frecuencia y sus representaciones gráficas no solo facilitan la organización de los datos, sino que son en sí una medida descriptiva de la distribución de frecuencia que tienen las variables del clima.

4.2.2. MEDIDAS NUMÉRICAS DESCRIPTIVAS

El tipo de análisis estadístico gráfico que se ha realizado, se complementa con el de tipo numérico, pues es más preciso y útil, interesándonos cuatro propiedades básicas que caracterizan a las variables:

- ◆ La localización del centro de la distribución, es decir, las *medidas de tendencia central*. Siendo las principales la media, la mediana y la moda.
- ◆ Qué tanto tienden los valores individuales a desviarse de la medida de tendencia central (grado de *variación o dispersión*). El software estadístico, comúnmente incluye un tipo de gráfica que nos ilustra claramente los conceptos que en la tabla IX se nombran como amplitud y amplitud intercuartil, se llama Box & Whisker y utilizando los datos de la temperatura exterior del año 1998 tenemos el siguiente cuadro. Las líneas indican la amplitud y el cuadro del centro indica la amplitud entre el cuartil $Q_{.25}$ (también conocido como el Q_1) y el cuartil $Q_{.75}$ (ó Q_3). Y también se muestra la relación que guardan las medidas de desviación con la mediana, que es una de las medidas de tendencia central.

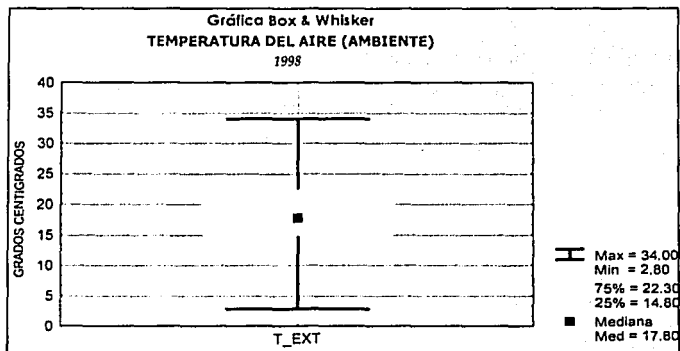


FIGURA 45. Gráfica que muestra amplitud y amplitud intercuartil

- ◆ El grado de asimetría, o sea, la falta de simetría entre un lado y otro del punto en el que se encuentra la mayor densidad de frecuencia. Ejemplificando este concepto por medio de una ilustración, tomamos los polígonos de frecuencia que representan la distribución de la Humedad exterior para los meses de febrero y octubre de 1997, los que como se aprecia tienen asimetrías diferentes. También se muestra la relación que guarda este concepto con las medidas de tendencia central.

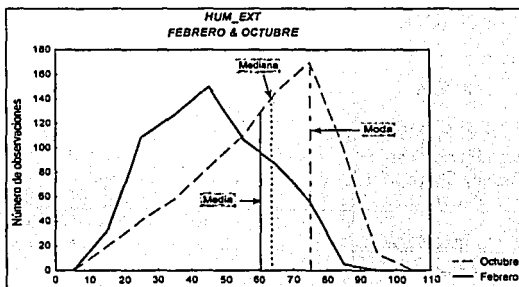


FIGURA 46. Distribuciones asimétricas. Humedad exterior.

- ◆ “El grado de variación, o la velocidad con que sube y baja la distribución de izquierda a derecha” Chou (1972), en otras palabras, qué tan puntiaguda es la distribución de frecuencias. (Medida de agudeza).

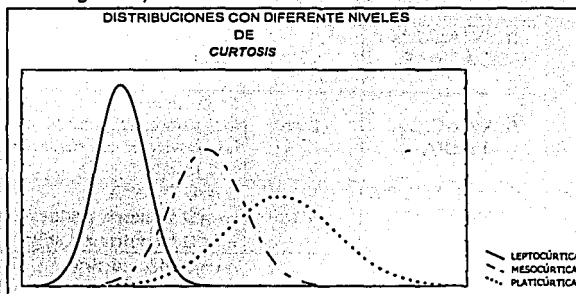


FIGURA 47. Variaciones de curtosis

Veamos en el cuadro sinóptico cuáles son estas medidas (estadísticas) más comunes:

TABLA XIV. *Tabla de medidas numéricas descriptivas*

Nombre	Símbolo	Descripción	Fórmula ²⁶
Medidas de tendencia central	MEDIA	\bar{X}	La suma de los valores individuales de una serie, dividida por el total de elementos en la muestra $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$
	MEDIANA	X_s	El valor que divide a la mitad una serie ordenada por magnitud
	MODA	X_m	El valor que aparece con más frecuencia en una serie
Medidas de dispersión ó desviación	AMPLITUD (también se llama Rango, Oscilación o Recorrido)		La diferencia entre el valor máximo y mínimo en una muestra.
	AMPLITUD INTERCUARTIL ²⁷	$q_{.25} - q_{.75}$	Analiza la situación del intermedio de la distribución. Incluye la mitad (central) de las observaciones. $q_{.25} = (n+1)/4$ $q_{.50} = (n+1)/2$ $q_{.75} = (3n+3)/4$
	VARIANZA	S^2	Es la media aritmética de los cuadrados de las desviaciones respecto a la media aritmética $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$
	DESVIACIÓN ESTÁNDAR	S	La raíz cuadrada de la varianza $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2}$
	DESVIACIÓN MEDIA	DM	Es la media aritmética de los valores absolutos de las desviaciones respecto a la media $DM = \frac{\sum_{i=1}^n x_i - \bar{X} }{n}$
Medidas de Asimetría y Agudeza ²⁸	COEFICIENTE DE ASIMETRÍA	α_3	$\alpha_3 = 0$ Simetría $\alpha_3 > 0$ asimetría positiva $\alpha_3 < 0$ asimetría negativa
	COEFICIENTE DE CURTOSIS	α_4	$\alpha_4 = 0$ Mesocúrtica $\alpha_4 > 0$ Leptocúrtica $\alpha_4 < 0$ Platicúrtica

²⁶ Estas fórmulas solo aplican para datos desagrupados y provenientes de una muestra.

²⁷ Los valores cuantiles utilizados comúnmente son los percentiles, deciles y cuartiles; son los valores que dividen la distribución de probabilidad en: percentiles: 100 intervalos, cada uno con probabilidad 0.01, Deciles: 10 intervalos cada uno con probabilidad 0.1, cuartiles: 4 intervalos con probabilidad 0.25 cada uno.

²⁸ Se consideran las fórmulas que Statística utiliza basadas en el concepto de momento respecto a la media.

En la tabla anterior se introdujo el símbolo de sumatoria que simplemente quiere decir "sumar los valores de X empezando con el primer valor $i = 1$ y terminando con el valor $i = n$ ", debido a que la parte de estadística descriptiva esta dedicada principalmente a los estudiantes no se explican a detalle las medidas de asimetría y curtosis pues involucran el concepto de momentos (se puede consultar a Portus (1985)), además de que existen diferentes coeficientes para el cálculo de estas medidas.

Por sus características matemáticas, la media y la varianza son las principales medidas de tendencia central y dispersión.

La obtención de estas estadísticas descriptivas utilizando un software como *Statistica* es sumamente fácil y existen varias maneras de hacerlo, esto aligera mucho la labor cotidiana de quien tiene que analizar datos climatológicos, como parte de este trabajo, se creó un macro que es ejecutado en el módulo de estadísticas básicas mediante la combinación de las teclas Ctrl + F para este fin, puede utilizarse en cualquiera de los 24 archivos mensuales y la información que se genera además de ser un primer acercamiento con las estadísticas de la muestra horaria, también permite la creación de un reporte mes por mes para 1997 y 1998 (puede verse al final del ANEXO C); estos cuadros pueden imprimirse desde el programa estadístico o editarse en un procesador de texto "copiando y pegando" previamente.

Para estudiar y clasificar el clima de un lugar, la temperatura del aire (T_{EXT}) es el elemento climatológico más importante, la metodología del trabajo climatológico requiere calcular la Temperatura Media Diaria (TMD), en nuestro caso, con base en las 24 observaciones horarias; para después con las 30 ó 31 TMDs de todo un mes obtener la Temperatura Media Mensual (TMM) y finalmente tomando las 12 TMMs de un año se realiza el cálculo de la Temperatura Media Anual (TMA); como vemos el uso de la estadística \bar{x} es muy útil en el trabajo meteorológico.

La OMM(1990) recomienda que se contemple, dentro de las publicaciones que se elaboren en una estación meteorológica, un Informe llamado Volumen de Temperatura en el cual se deben incluir las medias de:

- la temperatura ambiente (T_{EXT}), es decir la TMD
- la temperatura máxima (T_{MAX})
- la temperatura mínima (T_{MIN})

y los valores extremos de la temperatura ambiente, esto mensualmente.

En la estación meteorológica se publican gráficas y tablas de la TMD correspondientes a uno o dos meses pero el software con que se generan (Weatherlink), no permite visualizar los datos de la gráfica, por lo que se realiza este reporte manualmente. Para auxiliar en esta labor se elaboró un macro.

Solo se requiere estar en el módulo *Basis Statistics and Tables* y correr el macro con la combinación de teclas Ctrl+E, para después copiar la tabla y pegarla en un editor de texto como Word y darle formato final. Se crearon las tablas correspondientes a cada mes de 1997 y 1998, acompañado de sus respectivas gráficas (ver ANEXO B), así como los reportes anuales.

4.2.3 ANÁLISIS BIVARIADO

Muchos problemas estadísticos surgen de la necesidad de estudiar la relación entre dos o más variables y en el estudio del clima no se da la excepción. Enseguida se explican de manera práctica solo dos de las principales técnicas para estudiar simultáneamente dos variables y encontrar la forma en que se encuentran interrelacionadas.

4.2.3.1 TABLAS DE CONTINGENCIA

En muchas ocasiones se necesita determinar si hay enlace entre dos características o rasgos diferentes en los que ha sido clasificado un elemento de nuestro interés, por ejemplo, para estudiar el comportamiento general del viento (que en términos físicos es un vector, pues tiene una magnitud y dirección), se utiliza un tipo de cuadro estadístico sencillo y de mucha utilidad para encontrar o verificar las relaciones que existen entre ambas variables.

Este tipo de cuadro se llama *Tabla de Contingencia*, en donde se relacionan dos o más características, y estas a su vez se subdividen en otras subcategorías. Estas tablas han servido para contestar a preguntas como ¿existe una conexión entre fumar tabaco y la predisposición a desarrollar cáncer pulmonar?.

Al estudiar el viento, nos interesa verificar una característica que es la DIRECCIÓN y otra que es la VELOCIDAD. Cuando solo trabajamos con dos rasgos, las tablas de contingencia también son llamadas tablas de dos sentidos o de doble entrada, pero pueden aplicarse a un mayor número de variables.

Recordemos que la DIRECCIÓN DEL VIENTO en las estaciones del bachillerato puede tomar cualquier valor de los 16 rumbos en la rosa náutica, mientras que para la VELOCIDAD DEL VIENTO se puede hacer una clasificación arbitraria o utilizar como ejemplo la escala Beaufort (OMM, 1979C) que además nos permite estimar la velocidad por la simple observación. Los rumbos de la rosa náutica, y las clases de viento pueden ser consideradas como las subcategorías respectivas.

Para la elaboración de una tabla de contingencia se siguen los mismos conceptos de clasificación y agrupación de datos que se emplearon en la sección 4.2.1 (pueden verse las figuras 40 y 42 que se refieren a las variables que nos ocupan), pero ahora se estudian las dos características simultáneamente; consideremos los datos del mes de agosto de 1998 para ejemplificar y tomando como subcategorías la clasificación de vientos según la escala Beaufort. Utilizando el archivo HO898.sta (o cualquier otro que se desee) en el módulo de estadísticas básicas y con la combinación de teclas Ctrl + Z se realiza un macro que crea una tabla de contingencia como la siguiente:

TABLA XV. Tabla de contingencia (Tipos de viento & dirección del viento)

AGOSTO DE 1998

VELOCIDAD	DIRECCIONES																
	N	NNE	EN	ENE	E	ESE	SE	SSE	S	SSW	SW	WSW	W	WNW	NW	NNW	TOTAL
CALMA	3	3	4	5	2	1	4	4	3	0	3	3	9	0	2	0	46
VENTOLINA	58	31	36	24	18	25	22	11	9	25	31	42	24	10	20	13	399
BRISA MUY DÉBIL	75	12	4	2	6	5	6	7	2	1	13	18	6	1	19	25	202
BRISA DÉBIL	64	7	0	1	0	1	2	1	0	0	0	2	0	0	4	6	88
BRISA MODERADA	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
TOTAL	204	53	44	32	26	32	34	23	14	26	47	65	39	11	45	44	739

Fuerza Beaufort	Nombre	Equivalencia de la velocidad a una altura tipo de 10 metros sobre terreno llano M/s
0	Calma	0.0 - 0.2
1	Ventolina	0.3 - 1.5
2	Brisa muy débil	1.6 - 3.3
3	Brisa débil	3.4 - 5.4
4	Brisa moderada	5.5 - 7.9
5	Brisa fresca	8.0 - 10.7
6	Viento fresco	10.8 - 13.8
7	Viento fuerte	13.9 - 17.1
8	Viento Duro	17.2 - 20.7

En la climatología estas tablas se utilizan para hacer un dibujo conocido como la rosa de los vientos que sirve para reportar semanal, mensual o anualmente, la dirección, frecuencia y velocidad de los vientos dominantes, en ese lapso de tiempo; existen variaciones sobre el mismo gráfico pero aquí se muestra uno de ellos, en donde la flecha indica que el viento dominante (siguiendo con los datos de agosto de 1998) provino principalmente del norte (cosa que podemos corroborar en la tabla anterior) y representó casi el 30% del total de 739 observaciones realizadas para estas variables.

DIRECCIÓN DEL VIENTO DOMINANTE: NORTE
 VELOCIDAD DOMINANTE: ENTRE 1.6 Y 3.3 m/s (BRISA MUY DÉBIL)

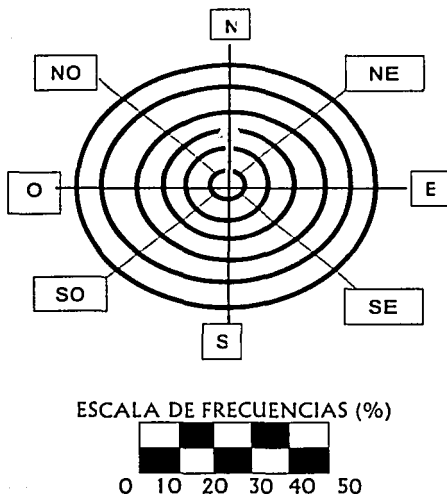


FIGURA 48. Rosa de los vientos (Agosto 1998)

También podemos observar que los vientos clasificados como ventolina son los que se presentaron con mayor frecuencia (399) durante dicho mes.

4.2.3.2. GRÁFICAS DE DISPERSIÓN

Los meteorólogos han estudiado las relaciones que existen entre diferentes elementos del clima y una de las que se ha podido probar es la que hay entre la humedad relativa y la temperatura del aire (llamada también temperatura ambiente).

Aún cuando existen diversos métodos de la estadística al estudiar la relación entre dos o más elementos del clima, generalmente se inicia con un análisis gráfico sencillo y muy útil para conocer la "forma" de esta unión; pues marcando los datos disponibles como puntos en el eje XY, se construye una *Gráfica de dispersión*, y al observarla se puede mostrar que no existe ninguna relación entre los elementos graficados, ó que la relación puede ser lineal o curvilínea.

Las investigaciones de muchos años anteriores y que están documentadas, indican que sí existe la relación temperatura del aire – humedad relativa la cual es curvilínea, sin embargo para períodos cortos de tiempo se presenta como lineal. Esto lo podemos confirmar empíricamente, es decir, con datos tomados en la actividad cotidiana de una estación meteorológica del bachillerato.

Tomemos para ejemplificar, los datos del 2 de febrero de 1997. Y vamos a establecer arbitrariamente el orden en que se colocan los nombres en el eje cartesiano, en este caso colcaremos en el eje horizontal los datos correspondientes a la temperatura del aire, mientras que en el eje vertical graficaremos los datos de humedad relativa.

Tengamos siempre presente que al realizar una gráfica, cualquiera que esta sea se debe especificar en ambos ejes lo que se esta analizando y en que unidades se maneja la escala de valores.

HORA	TEMPERATURA AMBIENTE (°C)	HUMEDAD RELATIVA (%)
1	13	52
2	12.9	51
3	11.9	55
4	11.7	55
5	11.1	58
6	10.7	58
7	10.2	60
8	12.5	54
9	14.9	46
10	17.2	39
11	18.4	34
12	19.2	32
13	20.4	28
14	21.7	25
15	22.2	25
16	22.1	27
17	22	24
18	21.2	27
19	18.8	39
20	18.1	40
21	17.3	41
22	16.3	40
23	15.6	43
24	14.6	45

Al ver el siguiente diagrama de dispersión se verifica que existe una tendencia de los valores bajos de temperatura (x) a asociarse con valores altos de humedad (y) y también vemos que el conjunto total de puntos se encuentra concentrado dentro de una zona angosta formando aunque de manera vaga, una línea recta.

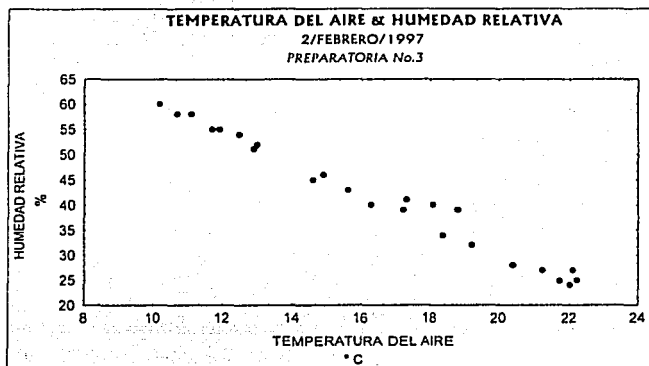


FIGURA 49. Diagrama de dispersión

En este caso la manera en que están colocados los puntos nos indica una cierta tendencia acerca de las dos variables, pues existe una propensión de los puntos de la gráfica de dispersión a agruparse a los lados de una línea recta.

Pero la estadística también nos permite medir cuantitativa o numéricamente el grado de relación entre x e y, mediante el *coeficiente de correlación lineal (producto-momento)*, dicha medida se caracteriza por no ser dependiente ni de la escala de medición ni del punto de origen en ella, y está establecido como:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

Donde s_x y s_y son las dos desviaciones estándar de muestra correspondientes a las variables x e y. Cabe hacer notar que existe otro coeficiente de correlación para variables de tipo ordinal, r_s llamado coeficiente de correlación por rangos de Spearman, debido a que no contamos con datos de ese tipo no se dará más detalle.

El valor de r varía entre los límites +1 y -1 y su magnitud indica el grado de asociación entre las variables. Un valor de $r=0$ señala que no existe relación entre las

variables, mientras que los valores $r = \pm 1$ son indicadores de una correlación perfecta positiva o negativa.

La Interpretación del coeficiente de correlación (r) debe hacerse teniendo siempre en mente que solo es una medida de covariabilidad mas no se puede pensar en una relación de causa-efecto, si aceptáramos esto llegaríamos a conclusiones tan equivocadas como el pensar que si aumentamos de peso, nuestra estatura será mayor (esto después de ver la alta correlación existente en un diagrama de dispersión de ambas variables).

Otro punto que debemos tener en cuenta es que esta medida de correlación es útil solamente si ambas variables están relacionadas linealmente, pues puede ocurrir que estén relacionadas de forma no lineal y obtendríamos un valor bajo de r , lo cual no pone en evidencia la carencia de covariación.

En la siguiente figura se muestran diferentes niveles de correlación lineal obtenidos al relacionar elementos del clima observados en la estación meteorológica de la ENP 3.

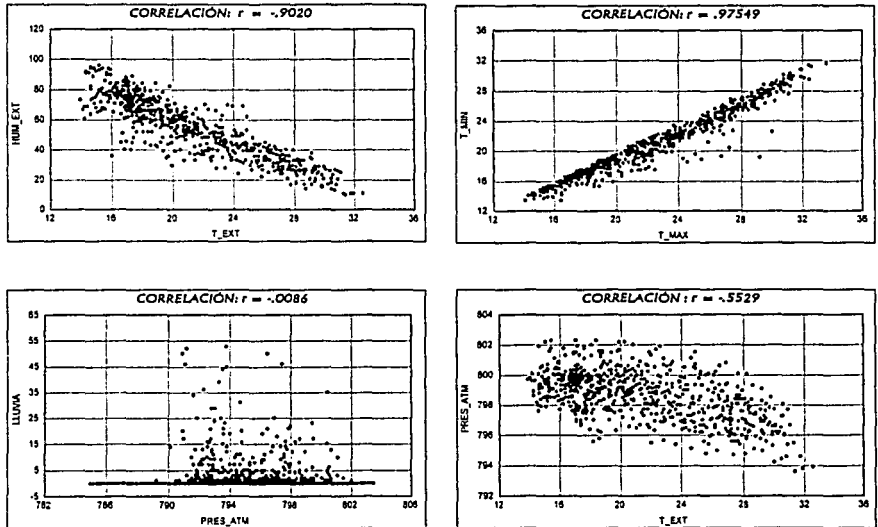


FIGURA 50. Diferentes niveles de correlación

En cuanto a la variación encontrada entre dos variables "puede deberse a causas que afectan a cada variable de una misma manera o de maneras opuestas, o

bien una de ellas es causa de la variación de la otra, o puede que la relación causal sea recíproca; todo esto es ajeno a la comprobación de la existencia de la correlación y del valor del coeficiente de correlación" Portus(1985).

En el caso de la estación se podría invitar al estudiante a encontrar los fundamentos climatológicos o físicos que explican las relaciones entre variables, incorporando de esta manera los conceptos de varias asignaturas y de manera incidental, pero estimulada por la inquietud o interés de un nuevo conocimiento.

Fue realizado un macro para ejemplificar la manera tan simple en que pueden crearse con *Statistica* los diagramas de dispersión y al mismo tiempo se informa del valor r , es decir, del nivel de correlación entre dos variables en este caso solo se puso como ejemplo a las variables de T_MAX y T_MIN (que aparecen en la figura anterior y utilizando el archivo HO698.STA). El macro se ejecuta con la combinación de teclas Ctrl + G.

4.3 ESTADÍSTICA INFERENCIAL

Hasta aquí hemos estudiado la parte descriptiva de la estadística que nos ha permitido organizar, analizar y presentar las características generales de la distribución muestral lo cual representa en el estudio del clima una base muy importante, sobre todo a partir de que con el uso de la computadora y el software estadístico se han ampliado las posibilidades de contar con un mayor volumen de datos y de herramientas para su análisis estadístico.

Sin embargo, recordemos que la estadística inferencial permite que con los datos muestrales que ahora tenemos podamos realizar estimaciones y pruebas de hipótesis con respecto a la población de origen, que en este caso es el clima de la zona y en específico de los elementos climatológicos que detecta Weather Monitor. Aunque aquí solo se mostrarán algunos usos de los datos del clima, no perdamos de vista que existe un gran número de posibilidades de estudio en lo que se refiere a la estadística inferencial sobre todo para quien gusta de realizar investigaciones específicas sobre el clima.

En esta sección del trabajo abordaremos la estadística inferencial y especialmente las pruebas de hipótesis y algunos ejemplos de aplicación, la dirigimos principalmente a los coordinadores de la estación meteorológica del bachillerato y a aquellos estudiantes que estén interesados en profundizar el estudio del clima desde un punto de vista estadístico, incluso se propone que aquellos que cursan la materia optativa de Probabilidad y Estadística en los últimos semestres del bachillerato (ya sea en la ENP o CCH), puedan realizar prácticas en la estación y/o ser parte del Club que se forma en cada plantel, lo que significaría impulsar a los futuros investigadores en campos multidisciplinarios.

Partimos del conocimiento o repaso de algunos conceptos necesarios para el planteamiento de hipótesis estadísticas como:

- 1) La diferencia entre una distribución empírica (muestral) y una teórica (modelo matemático que representa a la población), así como su relación.
- 2) Las características de la principal distribución teórica para el estudio de las variables de tipo continuo (pues a este tipo corresponden los datos que tenemos). Otros conceptos fundamentales como el teorema del límite central y la distribución t .
- 3) Un problema de estimación paramétrica, así como una alternativa no paramétrica a un caso planteado en las pruebas de hipótesis.

Se recomienda el apoyo de un libro de texto para aclarar cualquier duda en el desarrollo de este subtema. Se pueden usar los textos dedicados a la estadística a nivel bachillerato como son los de Hoel y Portus mencionados en las referencias bibliográficas, pues son sencillos e ilustrativos. Hay que tomar en cuenta que en adelante solo se considera el estudio de las variables continuas.

4.3.1 DISTRIBUCIONES DE PROBABILIDAD EMPÍRICAS Y TEÓRICAS (VARIABLES CONTINUAS)

La temperatura del aire (T_{EXT}) es la variable de mayor importancia en el estudio del clima, y la utilizaremos para el siguiente ejemplo. Si una persona consulta las tablas del Anexo B (el volumen de temperatura de 1997) y con base en los datos de las 12 temperaturas medias mensuales crea una tabla de frecuencias – estableciendo 5 clases –, y su correspondiente histograma, podrá observar el patrón de distribución para esa muestra mostrado en el inciso (a) de la siguiente figura. Pero ahora desea mayor detalle y entonces toma las temperaturas medias diarias para todo el año (349) y las clasifica en intervalos de clase más cortos, 11 en total, obteniendo un histograma como el del inciso (b), veamos que este histograma se acerca más a la curva en forma de campana que aparece ahí también. Decide seguir el proceso de aumentar el número de observaciones en la muestra y disminuir la amplitud de los intervalos de clase, esto lo hace consultando los datos horarios para todo el año que en total son 8291 y los agrupa en 20 clases creando un histograma (c) mas regular y que se aproxima todavía más a la curva que le acompaña. Continuando este proceso se llegaría a la *curva límite* que tiene forma de campana. Aquí se ilustraron frecuencias absolutas, pero se acostumbra trabajar frecuencias relativas para tener siempre áreas bajo la curva con valor unitario, esto por los conceptos de probabilidad.

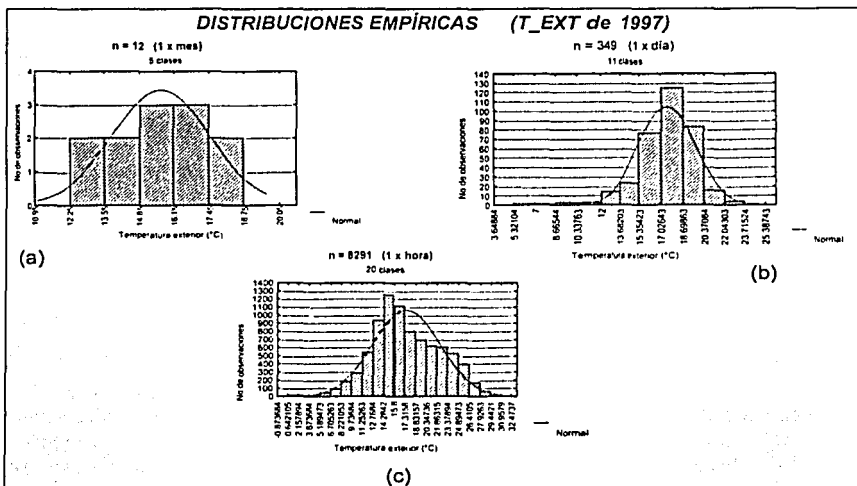


FIGURA 51. Relación entre distribución empírica (muestra) y teórica (población)

Debido a lo anterior se dice que "si el tamaño de la muestra es grande, puede esperarse que la distribución de frecuencias de la muestra sea una buena aproximación a la distribución de frecuencias de la población correspondiente a ella" Hoel (1988). Podemos ver ahora claramente que los histogramas muestran el comportamiento de la distribución de la muestra a la que llamaremos DISTRIBUCIÓN EMPÍRICA, la cual se aproxima cada vez más a la DISTRIBUCIÓN TEÓRICA, que es un modelo matemático de la distribución de la población, por cuyo medio se puede calcular la probabilidad de que una variable X se encuentre dentro de un intervalo específico de valores (como ejemplo podríamos calcular la probabilidad de que la TMD estuviera entre los 12 y los 15 grados centígrados).

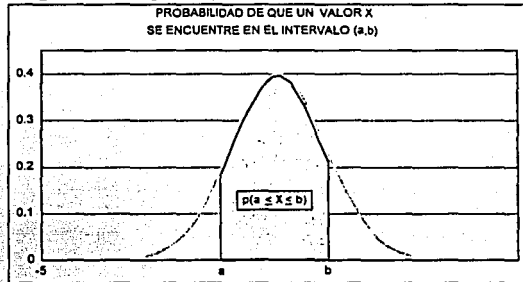


FIGURA 52. Representación gráfica de la probabilidad en variables continuas

Resumiendo (ver tabla XVI), cuando se estudian distribuciones de muestreo y sus modelos teóricos, se acostumbra nombrar a la distribución de muestreo DISTRIBUCIÓN EMPÍRICA que como sabemos se describe numéricamente por estadísticas, mientras que la DISTRIBUCIÓN TEÓRICA (que se refiere a la población) está definida por parámetros.

TABLA XVI. Diferencia entre distribución empírica y teórica

	DISTRIBUCIÓN EMPÍRICA	DISTRIBUCIÓN TEÓRICA
ORÍGEN	MUESTRA	Modelo matemático que describe una POBLACIÓN
CARACTERÍSTICAS QUE LA DESCRIBEN	ESTADÍSTICAS : \bar{x} s	PARÁMETROS: μ σ
TAMAÑO ($n < N$)	n	N

El símbolo μ corresponde a la media de la población y σ es la desviación estándar de la misma.

La importancia de las distribuciones teóricas es que están completamente definidas matemáticamente por fórmulas que aunque aquí no se detallan, nos abren la puerta para realizar inferencias.

4.3.2 DISTRIBUCIÓN NORMAL

El tipo de distribución teórica que se representa en la fig. 52, nos interesa principalmente por que (como lo vimos en la Tabla IV) los tipos de variables que trabajamos son continuas y la distribución normal es la más usada para estas variables. Es típica de muchas de las distribuciones que se encuentran en la naturaleza y la industria.

Su forma es simétrica, se desvanece a las orillas y se asemeja a una campana; como todas las distribuciones de probabilidad las curvas asociadas están definidas por una ecuación (que los matemáticos han estudiado ampliamente y que pueden ser consultadas en las referencias bibliográficas del tema), pero se puede considerar a la curva misma como una definición y así lo haremos en adelante.

Los parámetros que definen completamente la forma de esta curva normal son μ y σ , estos son los valores límites cuando la muestra es muy grande y correspondientes a \bar{X} la media muestral y a s la desviación estándar de la muestra. La curva tiene las siguientes características:

- 1) El área bajo la curva normal entre $\mu - \sigma$ y $\mu + \sigma$ es el 68% del área total, al 1% más cercano.
- 2) El área bajo la curva normal entre $\mu - 2\sigma$ y $\mu + 2\sigma$ es el 95% del área total, al 1% más cercano.
- 3) El área bajo la curva normal entre $\mu - 3\sigma$ y $\mu + 3\sigma$ es el 99.7% del área total hasta el 0.1% más próximo.

En la siguiente figura se muestran las proporciones que corresponden a cada área bajo la curva.

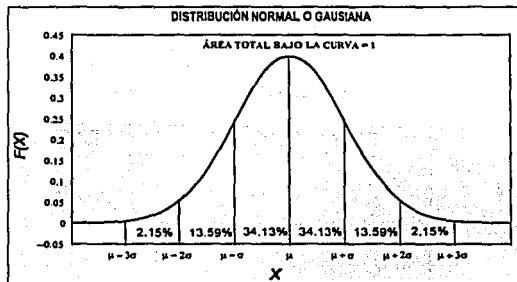


FIGURA 53. Propiedades de la curva normal

Aunque parece que no existe área bajo la curva más allá de 3 desviaciones estándar, la ecuación de la curva indica que esta se extiende de $-\infty$ a $+\infty$.

La distribución normal tiene toda una familia de curvas asociadas, esto es por que al cambiar los valores que toma alguno de sus parámetros, la forma de la curva cambia, a pesar de que el área bajo la curva siempre permanece constante con el valor unitario. Veamos la siguiente figura en donde solo varía el valor de tendencia central μ , permaneciendo constante el valor de σ .

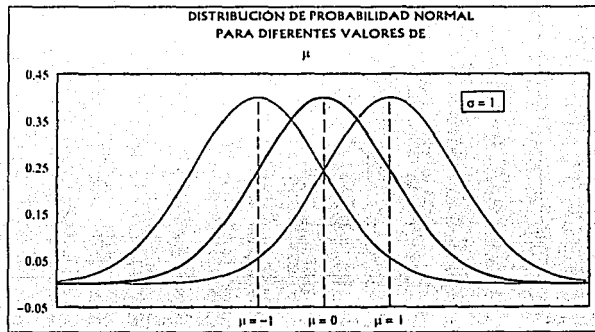


FIGURA 54. Curvas normales para diferentes valores de μ

Y cuando cambia la medida de dispersión pero μ permanece constante, se forman otras curvas

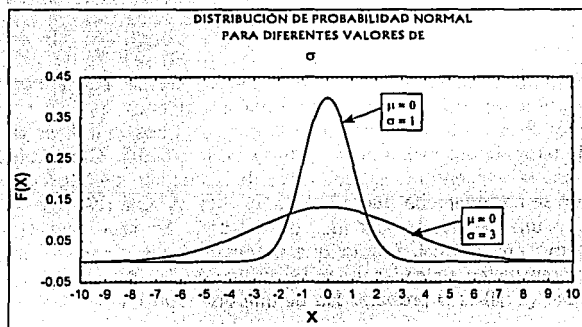


FIGURA 55. Curvas normales para diferentes valores de σ

Todo el grupo de curvas se puede reducir a una curva PATRÓN o ESTÁNDAR, que en lo que se refiere a cálculo, es la más sencilla pues tiene media igual a cero y desviación estándar uno. Enseguida se muestra gráficamente.

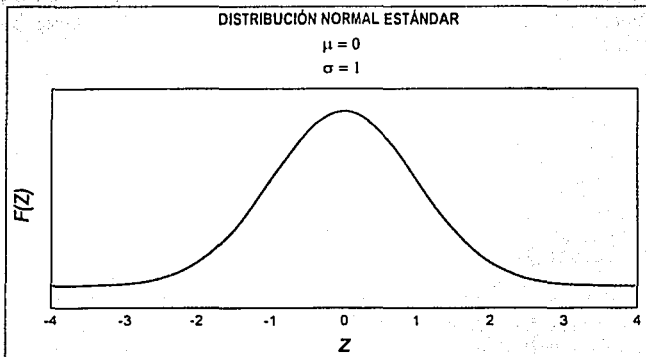


FIGURA 56. Curva normal estándar

De modo que a cualquier punto del eje horizontal en una curva normal, le corresponde un punto del eje Z de la curva normal estándar "este valor puede determinarse estableciendo la distancia (z) que hay entre el punto y la media de la curva, si se toma como unidad la desviación estándar" Hoel (1988). Esta relación se expresa en términos de X así:

$$z = \frac{X - \mu}{\sigma} \quad (1)$$

Esta fórmula nos permite encontrar el punto z de la curva normal estándar correspondiente al punto X en una curva normal cualquiera.

Veamos lo anterior mediante un ejemplo, digamos que X es la variable que representa la temperatura media anual (TMA) de la estación Gran Canal (propiedad del SMN y que se encuentra localizada cerca de la estación de la preparatoria 3). En dicha estación se ha observado durante 41 años esta variable registrándose una media de 16.9°C y una desviación estándar de 2.05°C. Se sabe que esta variable se distribuye como una normal, por lo que si tomamos en cuenta lo antes expuesto, podemos expresarla en unidades estándar utilizando la fórmula (1), pero además queremos saber ¿Cuál es la probabilidad de que una futura TMA sea menor o igual a 15°C?

Para comprender mejor este ejercicio usemos como apoyo la siguiente figura que nos muestra la relación entre la distribución teórica que sirve de modelo y la distribución normal estándar, así como la representación del concepto de probabilidad como una "área bajo la curva".

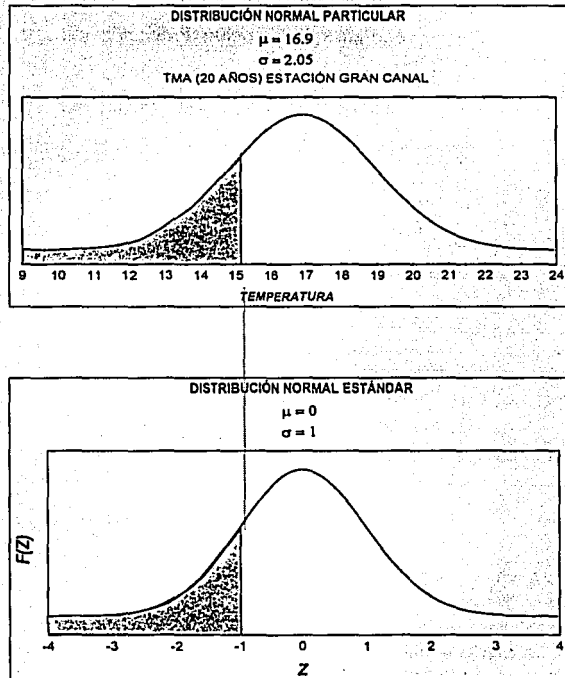


FIGURA 57. Correspondencia entre la distribución de X y Z
 Primero expresemos la interrogante en términos más cortos:

$$P(X \leq 15)$$

Ahora encontremos el valor Z que le corresponde al valor dado X aplicando la fórmula (1) y para simplificar aún más redondeamos los valores utilizados en los cálculos.

$$z = \frac{15 - 17}{2} = -1$$

expresando en probabilidades

$$P(X \leq 15) = P(Z \leq -1)$$

Y utilizando *Statística* (ver figura 29, opción *Probability calculator*) como señala la siguiente figura obtenemos

$$P(X \leq 15) = P(Z \leq -1) = 0.1587$$

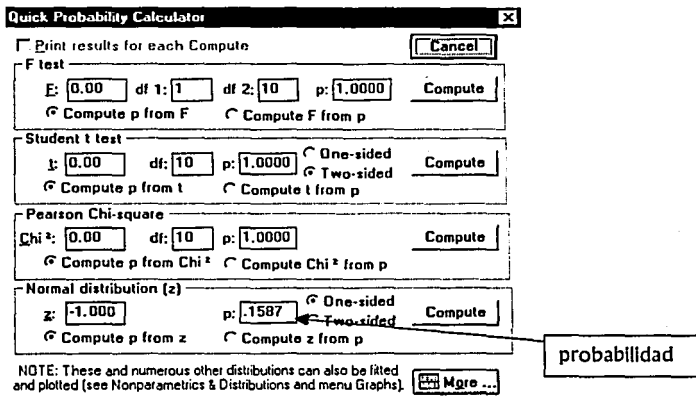


FIGURA 58. Módulo para el cálculo de probabilidades

Lo cual significa que hay casi un 16% de probabilidad de que se presente un año con temperatura media anual inferior o igual a 15°C en la estación Gran Canal.

Lo anterior ilustra la utilidad de la distribución normal estándar, y pensemos que con un razonamiento similar podemos obtener la probabilidad de que la TMA se encuentre en un intervalo específico, por ejemplo entre los 16 y 18 grados centígrados o cualquier otro intervalo deseado, recordando que el área bajo la curva es siempre 1.

Cabe hacer notar que *Statística* proporciona las probabilidades para el valor z en el intervalo $(-\infty, 0]$, en caso necesario hay que utilizar el concepto de simetría que implica cualquier distribución normal y en la mayoría de los textos de estadística se pueden consultar las tablas que nos indican el valor bajo la curva normal estándar, dado un valor z.

Recordemos que un modelo de población está definido por completo por ciertos parámetros (en el caso de la distribución normal por μ y σ) los cuales son

aproximados mediante los valores de muestra \bar{X} y s , por esto es importante conocer el teorema siguiente en el que se explica la forma en que se distribuye la media muestral.

4.3.3 TEOREMA DEL LÍMITE CENTRAL

El teorema del límite central nos dice que cuando el tamaño de la muestra n es grande ($n > 30$), la distribución de \bar{X} es aproximadamente normal (aún sin importar cuál sea el modelo de probabilidad a partir del que se obtuvo la muestra), teniendo una media μ y desviación estándar σ/\sqrt{n} por lo que la variable

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (2)$$

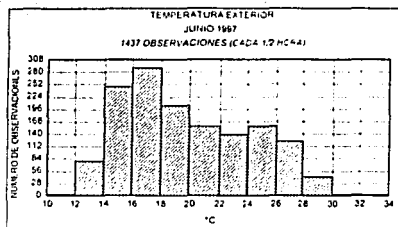
es utilizada para realizar inferencias respecto a μ cuando se conoce el valor de la varianza poblacional σ^2 .

En los casos en que no se conoce la varianza pero la muestra es grande, se permite utilizar la varianza muestral s^2 como estimación.

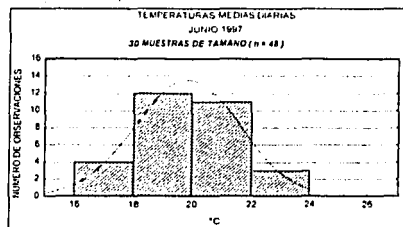
Para ilustrar consideremos los datos de TMD de junio de 1997, para ese mes se tienen 717 observaciones realizadas una cada hora, para calcular la temperatura media diaria se organizaron los datos en grupos de 24 observaciones que corresponden al total de un día y se obtuvo su media, por esto podemos decir que desde el punto de vista estadístico tenemos 30 muestras de tamaño $n=24$ para el mes.

Sin embargo por fortuna tenemos también un archivo que conserva las observaciones cada media hora. De modo que en dicho archivo el número de observaciones se incrementa a un total de 1437, o sea, que es una muestra más grande y al calcular ahora las correspondientes temperaturas medias diarias, se tienen 30 muestras con tamaño $n=48$ y de cada muestra se calcula la media. El inciso (a) de la siguiente figura muestra el histograma de las 1437 observaciones de T-EXT y observamos que se distribuye en forma parecida a una normal con simetría positiva, mientras que el inciso (b) muestra el histograma de las 30 temperaturas medias diarias, es decir, los 30 valores de la estadística \bar{X} , o medias muestrales.

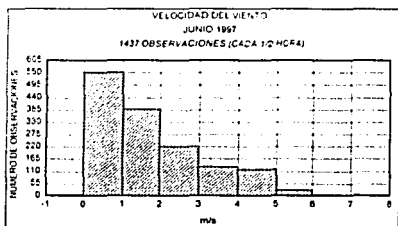
Si repetimos el mismo proceso pero ahora utilizando los datos de la velocidad del viento VEL_VTO observamos al clasificar y graficar los 1437 datos en un histograma que no se distribuyen como una normal sino en forma de J invertida (ver inciso (c)), sin embargo ahora al obtener las medias muestrales de los 30 días del mes y realizar un histograma de los valores de \bar{X} , vemos que también se distribuyen en forma de campana, recordemos que aquí $n=48$.



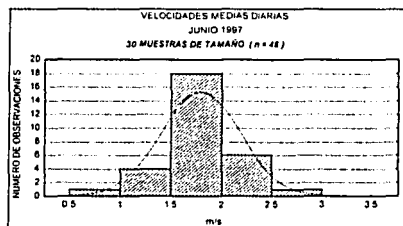
(a)



(b)



(c)



(d)

FIGURA 59. Distribuciones de los valores muestrales originales y de \bar{X}

Así que de forma visual comprobamos que el promedio muestral \bar{X} se distribuye en forma normal cuando n es grande (como en este caso $n > 30$) ya sea al tener como origen datos que se distribuyen en forma normal o no. Realizando el cambio de unidades de valores X a Z como lo hicimos en el tema anterior, transformamos la distribución a una estándar y a partir de ella se pueden realizar inferencias sobre el parámetro μ . Podemos ver que la distribución normal cobra mayor importancia debido a este teorema.

Véamos en seguida otro concepto que es de mucha utilidad en la estadística inferencial.

4.3.4 DISTRIBUCIÓN t DE STUDENT

Sucede frecuentemente que no se conoce el valor de la varianza de la población y además existe una muestra pequeña, lo cual impide formular inferencias respecto a μ por lo cual se debe hacer uso de otra distribución teórica que tiene mucha utilidad en la inferencia estadística por su semejanza con la distribución normal

estándar; aquí se introduce el concepto de *grados de libertad* definido como $v = n - 1$, conforme se tiene un número mayor de grados de libertad, la distribución t de Student tiende hacia la normal estándar; gráficamente se ve que los extremos de la primera son menos pronunciados que los de la segunda.

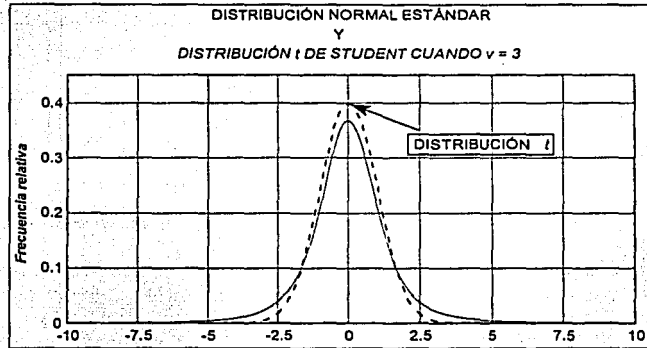


FIGURA 60. Comparación entre la distribución normal estándar y la t de Student

La variable t de Student es muy parecida a la fórmula 2, veamos que solo es sustituido el valor de σ por el de la desviación estándar de la muestra s .

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (3)$$

De modo que esta distribución solo depende del valor de n y como único requisito es necesario que la variable básica X posea una distribución normal.

4.3.5 ESTIMACIÓN DEL PARÁMETRO μ E INTERVALOS DE CONFIANZA

Cuando describimos las propiedades de la distribución normal mencionamos que el 95% del área bajo la curva se localiza entre dos desviaciones estándar alrededor de μ , esto lo podemos expresar de manera sencilla y en términos más exactos mediante esta expresión:

$$P \{ -1.96 < z < 1.96 \} = 0.95 \quad (4)$$

Y en forma gráfica lo anterior se representa así:

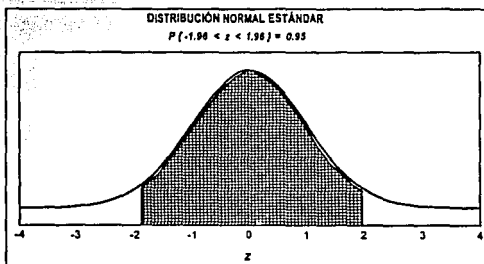


FIGURA 61. Intervalo que incluye el 95% del área bajo la curva normal estándar

Por otra parte, la estadística \bar{X} proveniente de una muestra nos permite hacer una aproximación del valor de la media de la población μ pero es conveniente especificar el grado de precisión de dicha estimación, esto es medido por los niveles de confianza, generalmente se utilizan el 95% o 90%. Veamos cómo se calculan.

Conocimos por el teorema del límite central que \bar{X} se distribuye en forma normal y que para expresarla en unidades estándar se utiliza la fórmula 2, de modo que la expresión en (4) es equivalente a

$$P \left\{ -1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96 \right\} = 0.95$$

Y resolviendo para μ tenemos

$$P \left\{ \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right\} \quad (5)$$

Esta fórmula sirve para obtener el *Intervalo de confianza* de 95% para la estimación de μ (en otras palabras el Intervalo de confianza nos indica qué tan creíble es la estimación). Si se desea un Intervalo del 90% solo basta sustituir el valor ± 1.96 por el número ± 1.64 . Como se puede ver en la fórmula es necesario saber el valor de σ sin embargo es poco común que esto suceda y en ese caso se utiliza la distribución t de Student.

En ese caso la fórmula a emplear sería

$$P \left\{ \bar{X} - t_{\alpha} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha} \frac{s}{\sqrt{n}} \right\} \quad (6)$$

donde t_{α} es el valor correspondiente de acuerdo a los grados de libertad $v=n-1$ que se requieran y al nivel de confianza deseado.

Veamos mediante un ejemplo cómo los conceptos revisados hasta ahora nos permiten resolver un problema práctico. Tenemos los datos de la estación San Juan de Aragón (la más cercana a la ENP 3) para el período 1971- 1980, es decir 10 años en los que se obtuvo una temperatura media anual de 16.4°C con una desviación estándar de 2.1°C. Consideramos que estos datos corresponden a una muestra con $n = 10$, $\bar{X} = 16.4$ y $s = 2.1$, y queremos saber con una certeza del 95% ¿Qué tan cerca se encuentra ésta estimación de la verdadera media de la población μ ?

Primero debemos recordar que la temperatura media anual se distribuye como una normal además el valor de σ no lo conocemos y la muestra es pequeña. por lo que cumple con los requisitos para utilizar la distribución t. Los grados de libertad aquí son $v=10-1=9$ y como el nivel de confianza que necesitamos es de 95% usaremos la fórmula

$$P \left\{ \bar{X} - 2.26 \frac{s}{\sqrt{n}} < \mu < \bar{X} + 2.26 \frac{s}{\sqrt{n}} \right\}$$

Y sustituyendo los valores

$$P \left\{ 16.4 - 2.26 \frac{2.1}{\sqrt{10}} < \mu < 16.4 + 2.26 \frac{2.1}{\sqrt{10}} \right\}$$

Realizando los cálculos tenemos que

$$14.9 < \mu < 17.9$$

es el intervalo de 95% de confianza y los puntos extremos del intervalo, es decir, 14.9 y 17.9 son los *límites de confianza* para μ , este intervalo nos da una idea de la exactitud con que se estima el parámetro.

4.3.6 PRUEBAS DE HIPÓTESIS ESTADÍSTICAS

Las pruebas de hipótesis (así como la estimación) corresponden a la parte inferencial de la estadística, ellas dan un camino objetivo para formular inferencias sobre las características de una población con base en muestras aleatorias.

En una prueba de hipótesis estadística se realiza una afirmación respecto a una característica desconocida de una población y esencialmente la prueba nos da una regla que permite en forma sistemática (y con base en la probabilidad) tomar una decisión respecto a si dicha afirmación se encuentra apoyada por la evidencia experimental, que es obtenida a través de una muestra aleatoria.

Cuando se trabaja con poblaciones que tienen una distribución normal generalmente las pruebas de hipótesis se reducen a realizar pruebas sobre los parámetros de la media poblacional μ , la desviación estándar σ ó alguna proporción, a esto se debe que se les llame pruebas PARAMÉTRICAS; pero existen otras en que independientemente del tipo de distribución de origen se pueden aplicar inferencias, incluso sobre datos categóricos, a estas pruebas de hipótesis se les conoce como NO PARAMÉTRICAS.

Iniciamos mencionando lo concerniente a las pruebas paramétricas y posteriormente trataremos un ejemplo, como alternativa no paramétrica, para efectuar una prueba de hipótesis sobre la diferencia de dos medias.

Para ejemplificar de manera sencilla la manera en que se plantea una hipótesis y los posibles errores en que se puede incurrir, considere un proceso penal en que como es costumbre se piensa que "el acusado es inocente hasta que se demuestre lo contrario"; esta frase puede expresarse mediante una afirmación llamada HIPÓTESIS NULA (H_0).

H_0 : Acusado = Inocente

Esta hipótesis nula se debe considerar verdadera hasta que exista suficiente evidencia en contra. El fiscal tratara de reunir pruebas que apoyen la falsedad de la hipótesis nula, por lo que debe existir una HIPÓTESIS ALTERNATIVA, esta se representa por H_1 y representa alguna forma de negación de la hipótesis nula. Para este caso sería

H_1 : Acusado = Culpable

Al probar la hipótesis nula, esta se puede RECHAZAR o NO RECHAZAR, cada una de estas decisiones tiene una consecuencia con respecto al estado real del acusado, veamos el siguiente cuadro:

TABLA XVII. Tipos de error en una prueba de hipótesis

	VEREDICTO	H ₀ CIERTO EL ACUSADO ES REALMENTE INOCENTE	H ₁ CIERTO EL ACUSADO ES REALMENTE CULPABLE
Rechazar H ₀	CULPABLE	ERROR TIPO I	CORRECTO
No Rechazar H ₀	INOCENTE	CORRECTO	ERROR TIPO II

Al elegir cualquiera de las dos acciones se puede cometer un error o realmente actuar correctamente.

Rechazar H₀ cuando ésta es cierta constituye lo que se denomina un error de Tipo I, mientras que No Rechazar H₀ cuando es falsa se considera un error de Tipo II. Para este caso, el error de tipo I representaría culpar a un inocente, mientras que el error de tipo II se cometería al declarar inocente a quien es culpable. Como se puede ver es más grave incurrir en un error de tipo I que de tipo II. (Generalmente se sigue esta premisa en las pruebas de hipótesis estadística).

La posibilidad de cometer estos errores se mide con probabilidades, que obviamente tienen un valor entre 0 y 1.

ERROR DE TIPO I

$$P(\text{Rechazar } H_0 \mid H_0 \text{ es cierta}) = \alpha$$

(Se lee "la probabilidad de rechazar H₀, dado que H₀ es cierta, es igual a alfa")
A este valor α se le nombra como el nivel de significancia de la prueba de hipótesis.

ERROR DE TIPO II

$$P(\text{No Rechazar } H_0 \mid H_0 \text{ es falsa}) = \beta$$

(Se lee "la probabilidad de No rechazar H₀, dado que H₀ es falsa, es igual a beta")

Hay que notar que estas probabilidades α y β están condicionadas a asumir la hipótesis nula como cierta o falsa respectivamente.

En el enfoque clásico de las pruebas de hipótesis, la persona que realiza la prueba actúa como si fuera un fiscal (es decir, la parte acusadora) el cual debe encontrar todas las pruebas que permitan refutar claramente la hipótesis nula de inocencia. Si finalmente no se puede comprobar la culpabilidad, esto no implica que realmente sea cierta la inocencia, sino simplemente que no se ha podido probar que el acusado es culpable y en este caso se actúa como si la hipótesis nula fuera cierta y se debe liberar al acusado. En otras palabras, al plantear una prueba de hipótesis lo que realmente se quiere es concluir que la hipótesis alternativa es la correcta.

La decisión de rechazar o no rechazar una hipótesis nula se basa en una estadística apropiada que recibe el nombre de *estadística de prueba*, para algunos valores de esta estadística se debe rechazar la hipótesis nula, dichos valores constituyen la *región crítica* y su área corresponde al tamaño del error tipo I, es decir, la probabilidad α . En la estadística práctica generalmente se establece el valor $\alpha = 0.05$ ó $\alpha = 0.01$, los cuales emplearemos en adelante.

En los casos prácticos sobre pruebas de hipótesis se puede generalizar sobre tres variantes para plantearlas, lo cual implica un tipo de región crítica para cada una, esto está determinado por la hipótesis alternativa, veamos lo que sucede gráficamente cuando se realizan pruebas sobre el parámetro μ (lo mismo sucede con otros parámetros o características de la población), μ_0 indica el valor particular que se prueba:

CASO 1.

Se establece una hipótesis de este tipo cuando no se tiene información específica respecto a los valores alternativos de la media, en caso de que el valor que se prueba no sea el verdadero.

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

A esta prueba alternativa se le da el nombre de *bilateral*, se entenderá mejor al ver la siguiente gráfica y observar que en ambos extremos de la curva se encuentran regiones críticas

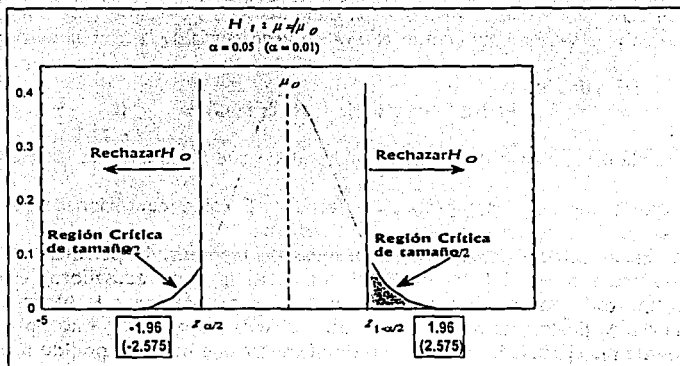


FIGURA 62. Región crítica bilateral, criterios de rechazo y valores críticos

Los valores señalados en un recuadro son los valores críticos cuando $\alpha=0.05$ y entre paréntesis están los que corresponden a $\alpha=0.01$.

Plantear una hipótesis de este tipo sería necesario para responder a interrogantes como la siguiente. Habíamos señalado que en la estación Gran Canal se observó en el período (1950-1990) una Temperatura Medía Anual (TMA) de 16.9°C , con una desviación estándar de 2.05°C . ¿Puede considerarse el comportamiento de la temperatura del aire durante 1997 en la preparatoria 3, como típico de esa zona?. Por lo que tendríamos que plantear una hipótesis del tipo que estamos describiendo y quedaría así:

$$H_0 : \mu = 16.9$$

$$H_1 : \mu \neq 16.9$$

más adelante se efectuarán las hipótesis que se van planteando.

En los siguientes dos casos se dice que la hipótesis alternativa es *unilateral*

CASO 2.

Para algunos problemas se tiene la certeza de que si la media no es igual al valor postulado bajo H_0 entonces su valor debe ser mayor.

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

En este caso la región crítica se ubica a la derecha de la curva de distribución.

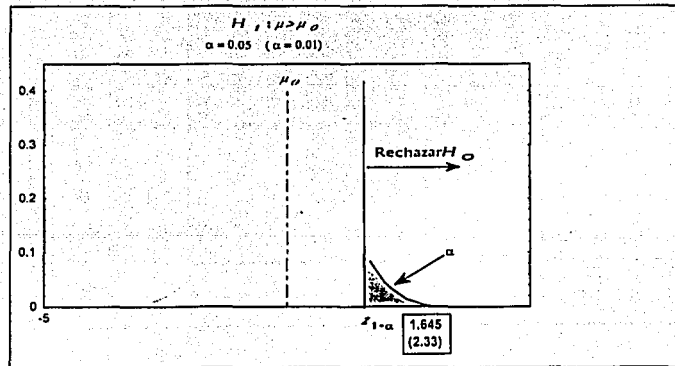


FIGURA 63. Región crítica unilateral (derecha) y criterio de rechazo

Veamos un ejemplo práctico. En los 41 años que se ha observado la temperatura del aire en la estación Gran Canal se encontró que 1982 fue el año con TMA más elevada con 18°C , pero en 1998 se presentó una temperatura que ahora es considerada por el SMN como la extrema histórica (33.9°C el 9 de mayo). Es obvio que fue un año caluroso, en la ENP 3 se registró una TMA de 18.07°C (ver Anexo C) pero ¿Se puede considerar más caluroso que 1982?. Para realizar esta prueba, la hipótesis nula y alternativa quedarían especificadas de la siguiente manera:

$$H_0: \mu = 18$$

$$H_1: \mu > 18$$

CASO 3.

Y en otros problemas se cree que si la media no es igual a μ_0 , entonces es menor

$$H_0: \mu = \mu_0$$

$$H_1: \mu < \mu_0$$

Y aquí, la región crítica está a la izquierda de la curva.

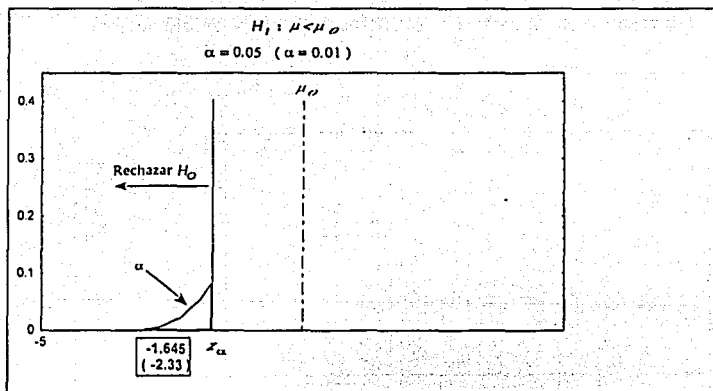


FIGURA 64. Región crítica unilateral (izquierda) y criterio de rechazo

Requerimos plantear una hipótesis de este tipo para resolver el siguiente problema. El Distrito Federal tiene en su gran mayoría un clima templado, pero en la zona noreste, donde se localiza la ENP 3 el clima es semiseco. En la estación Tacubaya (SMN) se ha registrado una humedad relativa media para el mes de agosto de 71% (1950-1980); considerando los datos de la ENP 3 (1997-1998) tenemos una muestra de $n=62$ con media de 64.5% y desviación estándar de 7 para la humedad de la zona y queremos probar si los datos empíricos permiten confirmar lo que teóricamente debe ser, pues a mayor temperatura menor humedad en la ciudad de México ¿Es menor la humedad en la zona noreste que en Tacubaya durante ese mes? Quedando las hipótesis así:

$$H_0 : \mu = 71$$

$$H_1 : \mu < 71$$

Estudiamos lo que sucede en la prueba de hipótesis respecto a una media, recordemos que es necesario trabajar con datos que puedan ser representados por una distribución teórica normal.

Generalmente cuando se realizan pruebas de hipótesis respecto a una media, se utiliza la estadística de prueba señalada con la fórmula 2. En la tabla XVIII se resumen los criterios para el rechazo de una hipótesis respecto a una media así como las estadísticas de prueba a utilizar y los valores críticos correspondientes.

En el punto 4.3.4 decíamos que en muchos casos prácticos no se conoce el valor de σ y entonces es necesario tomar la información que hace falta sustituyéndola por el valor s , es decir, la desviación estándar muestral, esto solo es permitido si la muestra es grande.

Si el tamaño de la muestra es pequeño ($n < 30$) y además se desconoce el valor σ , no es correcto utilizar la prueba especificada antes. Sin embargo, si la muestra proviene de una población normal (o con un grado razonable de aproximación), la prueba de hipótesis se puede realizar utilizando el estadístico de prueba:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

El cual es un valor de una variable aleatoria que cumple con una distribución t (Student) con $n - 1$ grados de libertad. Dependiendo del planteamiento de la hipótesis alternativa, también tendrá asociada una región crítica correspondiente (como lo vimos antes).

Lo relativo a las pruebas de hipótesis respecto a una media, se puede resumir en el siguiente cuadro:

TABLA XVIII. Criterios de rechazo (Prueba de una media)

PRUEBA DE HIPÓTESIS PARA UNA MEDIA CUANDO SE MUESTRAN DISTRIBUCIONES NORMALES		
HIPÓTESIS NULA $H_0: \mu = \mu_0$		
σ^2 conocida	Estadística De prueba	Criterios de rechazo
	$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	$H_1: \mu \neq \mu_0$ Rechazar H_0 cuando $z \leq z_{1-\alpha/2}$ ó $z \geq z_{1-\alpha/2}$ $H_1: \mu > \mu_0$ Rechazar H_0 cuando $z \geq z_{1-\alpha}$ $H_1: \mu < \mu_0$ Rechazar H_0 cuando $z \leq z_{\alpha}$
σ^2 desconocida	Estadística De prueba	Criterios de rechazo
	$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$	$H_1: \mu \neq \mu_0$ Rechazar H_0 cuando $t \leq t_{1-\alpha/2, n-1}$ ó $t \geq t_{1-\alpha/2, n-1}$ $H_1: \mu > \mu_0$ Rechazar H_0 cuando $t \geq t_{1-\alpha, n-1}$ $H_1: \mu < \mu_0$ Rechazar H_0 cuando $t \leq t_{\alpha, n-1}$

DIFERENCIA DE DOS MEDIAS

En muchos problemas estadísticos requerimos comparar dos medias provenientes de distribuciones normales, ilustremos con un ejemplo. Una persona estudia el clima del noreste del D.F, específicamente respecto a la temperatura, cuenta con datos de la estación Gran Canal y de San Juan de Aragón, y quiere saber si puede considerar las dos muestras como si provinieran de la misma población.

Para realizar pruebas de hipótesis respecto a la diferencia entre dos medias, se considera el siguiente teorema (el cuál no se define formalmente por estar dirigido a no-estadísticos):

Si x (con tamaño n_x) e y (con tamaño n_y) poseen distribuciones normales independientes, cuyas medias son μ_x y μ_y y cuyas desviaciones estándar son σ_x y σ_y , entonces la variable $\bar{x} - \bar{y}$ poseerá una distribución normal con media $\mu_x - \mu_y$ y una desviación estándar dada por la fórmula:

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\sigma_x^2 + \sigma_y^2} = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

y se plantea la hipótesis de valor nulo (pues supone que no hay diferencias) dada por

$$H_0: \mu_x - \mu_y = \delta_0$$

Se puede asignar el valor deseado a δ_0 especialmente en nuestro ejemplo tendría el valor de cero.

Como se realizó antes, en el cuadro que se presenta a continuación se resumen los criterios a utilizar en una prueba como la planteada, así como las estadísticas de prueba que deben utilizarse en caso de conocer o no el valor σ y los valores críticos:

TABLA XIX. Criterios de rechazo (prueba de dos medias)

PRUEBA DE HIPÓTESIS PARA DOS MEDIAS CUANDO SE MUESTRAN DISTRIBUCIONES NORMALES		
HIPÓTESIS NULA $H_0: \mu_X - \mu_Y = \delta_0$		
	Estadística de prueba	Criterios de rechazo
σ^2 conocidas	$z = \frac{\bar{x} - \bar{y} - \delta_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$	$H_1: \mu_X - \mu_Y = \delta$ Rechazar H_0 cuando $z \leq z_{\alpha/2}$ ó $z \geq z_{1-\alpha/2}$ $H_1: \mu_X - \mu_Y > \delta$ Rechazar H_0 cuando $z \geq z_{1-\alpha}$ $H_1: \mu_X - \mu_Y < \delta$ Rechazar H_0 cuando $z \leq z_{\alpha}$
	Estadística de prueba	Criterios de rechazo
σ^2 desconocidas	$t = \frac{\bar{x} - \bar{y} - \delta_0}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$	$H_1: \mu_X - \mu_Y = \delta$ Rechazar H_0 cuando $t \leq t_{\alpha/2, m}$ ó $t \geq t_{1-\alpha/2, m}$ $H_1: \mu_X - \mu_Y > \delta$ Rechazar H_0 cuando $t \geq t_{1-\alpha, m}$ $H_1: \mu_X - \mu_Y < \delta$ Rechazar H_0 cuando $t \leq t_{\alpha, m}$ donde $m = n_x + n_y - 2$

Podemos ver entonces que realizar una prueba de hipótesis consiste en seguir una regla para tomar una decisión. Esta regla generalmente se fundamenta en un valor muestral que suponemos similar a una variable x (distribución teórica). Dividimos los valores de la variable x en dos grupos: "en uno de ellos todos aquellos valores asociados con el rechazo de H_0 y los cuales forman la llamada región crítica de la prueba; en el otro grupo, todos los valores que se asocian con la "aceptación de H_0 " Hoel (1976).

No rechazar una hipótesis nula es una cuestión de decisión práctica y no implica que tenga que creerse que la afirmación hecha en ella sea precisamente correcta, mas bien significa que los datos de la muestra son compatibles con el valor postulado en la hipótesis.

Se esbozan de forma general las cinco etapas que se siguen al realizar una prueba de hipótesis estadística:

- 1) Formulación de una hipótesis nula simple (asignando un valor particular al parámetro de interés) y una hipótesis alternativa compuesta apropiada – que refleja el intervalo de valores que puede tomar el parámetro-, esta última se “acepta” cuando la hipótesis nula debe ser rechazada.
- 2) Especificar el tamaño máximo del error de tipo I (α), llamado también nivel de significancia (implica que la evidencia muestral es tal que garantiza el rechazo de H_0 a un nivel dado α). Generalmente se establece que $\alpha=0.05$ ó $\alpha=0.01$, pero la elección de este valor depende de las consecuencias de cometer un error de éste tipo. También se debe tener en cuenta que al disminuir el valor de α , se aumenta el de β (la probabilidad del error tipo II).
“Cuando la prueba de hipótesis produce un valor de muestra que se encuentra en la región crítica de la prueba, se dice que el resultado es *significativo*; en caso contrario, se dice que el resultado *no es significativo*” Hoel (1988).
- 3) Construir un criterio para probar la hipótesis nula contra la hipótesis alternativa, con base en la distribución muestral del estadístico de prueba, así como en lo planteado en la hipótesis alternativa.
- 4) Calcular con los datos disponibles, el valor del estadístico de prueba que es la base para la toma de decisión.
- 5) Decidir finalmente si se rechaza o no la hipótesis nula. Se rechaza H_0 en favor de H_1 cuando el valor de la estadística de prueba se encuentra dentro de la región crítica, y si éste valor no se encuentra dentro de dicha región crítica, no puede rechazarse la hipótesis nula. Y recordando siempre que el hecho de aceptar una hipótesis nula no significa que se probó como verdadera.

Para ilustrar con algunos ejemplos las pruebas de hipótesis, consideremos los problemas mencionados al estudiar los tipos de regiones críticas asociadas a las hipótesis alternativas que pueden ocurrir.

EJEMPLO 1.

En el caso 1 mencionamos que en la estación Gran Canal (considerando un período de 41 años) se tuvo un promedio de temperatura del aire de 16.9°C con una desviación estándar de 2.05°C .

La temperatura de un año se calcula con base en las medias de la temperatura mensual lo hace pensar (por el teorema del límite central) que los datos provienen de una población normal (recordemos que esto es un requisito para este tipo de prueba de hipótesis), nosotros podemos considerar que se obtuvo una muestra de tamaño $n=12$, correspondiente al año 1997 con media $\bar{x} = 17.4$ y $s=1.4$.

Para responder la pregunta acerca de si el año 1997 presentó un comportamiento que se pueda considerar como típico de la zona y siguiendo los pasos para realizar la prueba de hipótesis tendríamos que

- 1) Plantear las hipótesis correspondientes

$$H_0 : \mu = 16.9$$

$$H_1 : \mu \neq 16.9$$

- 2) Establecer el valor del error tipo I en $\alpha = 0.05$

La hipótesis alternativa es bilateral porque solo nos interesa saber si podemos considerar que la temperatura del aire (T_{EXT}) en la ENP 3 difiere de la observada en Gran Canal, sin considerar que haya sido inferior o superior al valor propuesto en la hipótesis nula. Como conocemos el valor σ emplearemos el estadístico de prueba z con $n=12$; para este nivel de significancia tendremos los valores críticos para la prueba bilateral de -1.96 y 1.96 .

- 3) El criterio es: Rechazo de la hipótesis nula si $z < -1.96$ ó $z > 1.96$, donde

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- 4) Cálculos:

$$z = \frac{17.4 - 16.9}{2.05 / \sqrt{12}} = 0.89$$

5) Decisión: como $z=0.89$ cae dentro del intervalo -1.96 a 1.96 la hipótesis nula no puede ser rechazada, en otras palabras, la diferencia entre $\bar{y} = 17.43$ y $\mu = 16.9$ puede ser atribuida al azar.

Podemos concluir que la evidencia muestral no permite rechazar la idea de que la temperatura media anual de 1997 haya tenido un comportamiento típico de acuerdo a los años observados anteriormente.

Ejemplo 2.

Cuando ilustramos el tipo de región crítica en el caso 2, presentamos los datos y la pregunta siguiente:

Estación Gran Canal (1950-1990): 18.0° C de Temperatura Media Anual durante el año más caluroso (1982) ¿la temperatura de 1998 tuvo un comportamiento semejante al de 1982 ó realmente lo superó como indicaban los medios de comunicación?

Fundamentamos la existencia de una distribución normal en el teorema del límite central (como en el ejemplo anterior) y también en las fuentes bibliográficas del tema climatológico (OMM, 1990). Consideramos que durante 1998 tenemos una muestra de tamaño $n=12$ cuya media muestral fue de 18.07° C y desviación estándar de 2.73° C.

1) Para realizar esta prueba, la hipótesis nula y alternativa quedarían especificadas de la siguiente manera:

$$H_0 : \mu = 18$$

$$H_1 : \mu > 18$$

2) Si especificamos un nivel de significancia de $\alpha = 0.01$

3) El criterio es: Como en este caso desconocemos el valor de σ y como la muestra es pequeña no sería adecuado sustituir su valor por el de s , por lo que debemos utilizar la estadística de prueba correspondiente a la distribución t con 11 grados de libertad. Rechazo de la hipótesis nula si $t_{11} > 2.718$, donde

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

4) Cálculos:

$$Z = \frac{18.07 - 18.0}{2.73/\sqrt{12}} = 0.0888$$

5) Decisión: como $t = 0.0888 < t_{11} = 2.718$ no puede rechazarse la hipótesis nula.

Podemos concluir que la evidencia muestral no permite confiar en la afirmación de que 1998 fue un año más caluroso que los anteriores.

Ejemplo 3.

El Distrito Federal tiene en su gran mayoría un clima templado, pero en la zona noreste, donde se localiza la ENP 3 el clima es semiseco. En la estación Tacubaya (SMN) se ha registrado una humedad relativa media para el mes de agosto de 71% (1950-1980); considerando los datos de la ENP 3 (1997-1998) tenemos una muestra de $n=62$ con media de 64.5% y desviación estándar de 7 para la humedad de la zona y queremos probar si los datos empíricos permiten confirmar lo que teóricamente debe ser, pues a mayor temperatura menor humedad ¿Es menor la humedad en la zona noreste que en Tacubaya durante ese mes?

1) Quedando las hipótesis nula y alternativa así:

$$H_0 : \mu = 71$$

$$H_1 : \mu < 71$$

2) Especificando un nivel de significancia de $\alpha = 0.01$

3) Criterio: En Este caso no tenemos el valor de σ , pero el tamaño de la muestra nos permite sustituirlo por la desviación estándar de la muestra por lo que se rechaza de la hipótesis nula si $z < -2.33$

4) Cálculos

$$Z = \frac{64.5 - 71}{7/\sqrt{62}} = -7.31$$

5) Decisión: como el valor obtenido de z se encuentra en la región crítica pues es $-7.31 < -2.33$ se rechaza la hipótesis nula y podemos considerar para fines prácticos que la humedad para la zona noreste del D.F es menor que la de Tacubaya como se esperaba por la propia teoría climatológica.

4.3.7 PRUEBAS NO PARAMÉTRICAS

Los métodos estadísticos no paramétricos son procedimientos inferenciales que generalmente no se encuentran sujetos a la forma de la distribución de la población de interés y también se usan comúnmente cuando se analizan datos de tipo categórico.

Aquí solo se presenta una prueba no paramétrica como alternativa a la prueba de hipótesis respecto a dos medias de la cual se hizo mención antes.

PRUEBA PARA COMPARAR DOS MEDIAS

Prueba U de Mann-Whitney

Para realizar esta prueba solo se requiere que las distribuciones de interés X_1, X_2, \dots, X_n , e Y_1, Y_2, \dots, Y_n , sean continuas.

Si se tiene una prueba de hipótesis en la que se comparan dos distribuciones continuas y en la que la hipótesis nula es:

$$H_0 : \mu_x = \mu_y$$

la hipótesis alternativa puede ser unilateral o bilateral, esta última solo indica que las distribuciones son diferentes y solo implica un desplazamiento en la tendencia central de una con respecto a la otra y no hace referencia a la dispersión. En otras palabras se supone que las distribuciones tienen la misma forma y dispersión.

La prueba U de Mann-Whitney es el equivalente a la prueba t de Student. Este procedimiento para probar la hipótesis de igualdad entre dos muestras se basa en:

- 1) La combinación de las n_x y n_y observaciones para formar un solo conjunto arreglado en orden creciente de magnitud.
- 2) La asignación de un RANGO a cada observación, comenzando con el rango 1 y finalizando con el rango $n_x + n_y$; cuando dos o más observaciones se encuentran repetidas, se asigna el promedio de los rangos que corresponderían a dichas observaciones si no estuvieran empatadas.
- 3) Se obtiene la suma de los rangos asociados a las observaciones de una muestra que se elige arbitrariamente, a la suma se le denota por R_1
- 4) El cálculo del valor de la prueba dada por

$$U = n_x n_y + \frac{n_x(n_x + 1)}{2} - R_1$$

Como se ve, el valor de U está en función de R_1 y del tamaño de las muestras.

- 5) Como se ha determinado que la distribución de U (cuando los tamaños de las muestras son mayores a 10) se aproxima a una distribución normal con una media

definida como:

$$E(U) = \frac{n_x n_y}{2}$$

Y una varianza dada por:

$$Var(U) = \frac{n_x n_y (n_x + n_y + 1)}{12}$$

Entonces bajo la hipótesis nula se tiene una variable aleatoria definida por:

$$Z = \frac{U - E(U)}{\sqrt{Var(U)}}$$

que es aproximadamente normal con media cero y desviación estándar uno; para muestras grandes.

Por esto, se calculan los valores de la media y la varianza para realizar una aproximación normal y tomar la decisión de rechazar o no la hipótesis, con base en el nivel de significancia.

En caso de tener una hipótesis alternativa bilateral y un valor muy grande o muy pequeño de U , es probable que se rechace H_0 , lo que a su vez ocurre al existir un valor grande o pequeño de R_1 respectivamente.

Veamos un ejemplo concreto para ilustrar claramente lo que se ha expuesto hasta ahora. Esta es una alternativa para probar un problema similar al expuesto en el ejemplo 2, solo que ahora requerimos considerar las 12 observaciones correspondientes a cada año que va a ser comparado. Recordemos que se sospecha que la temperatura media anual (TMA) correspondiente a 1998 es mayor que la de cualquier año anterior pero no pudimos (con base en la evidencia muestral) confirmar esta afirmación. Ahora vamos a comparar las temperaturas medias anuales (TMA) de 1998 en la ENP 3 y la correspondiente a la estación Gran Canal durante el período de observación de 1950-1990.

Las temperaturas medias mensuales de los dos períodos son las siguientes:

TEMPERATURAS MEDIAS MENSUALES												
ESTACIÓN	ENE	FEB	MZO	ABR	MYO	JUN	JUL	AGS	SEP	OCT	NOV	DIC
Gran Canal (1950-1990)	13.5	14.9	17.0	18.6	19.3	19.1	18.2	18.4	18.1	17.0	15.4	13.8
ENP 3 (1988)	14.3	15.6	18.9	22.0	22.9	21.4	19.3	18.5	18.1	16.6	17.4	15.2

¿Existe alguna razón para considerar que la temperatura media anual de 1998 en la ENP 3 difiere del comportamiento típico de la zona?

Consideramos una prueba de hipótesis de valor nulo:

$$H_0 : \mu_x = \mu_y$$

Donde x corresponderá a la estación Gran Canal, mientras que y se asigna a la ENP 3.

1) Y 2) Combinando las dos muestras, ordenando las 24 temperaturas medias y asignando el rango correspondiente tenemos que:

ESTACION	TMM	RANGO
GRAN CANAL	13.5	1
GRAN CANAL	13.8	2
ENP 3	14.3	3
GRAN CANAL	14.9	4
ENP 3	15.2	5
GRAN CANAL	15.4	6
ENP 3	15.6	7
ENP 3	16.6	8
GRAN CANAL	17.0	9.5
GRAN CANAL	17.0	9.5
ENP 3	17.4	11
ENP 3	18.1	12.5
GRAN CANAL	18.1	12.5
GRAN CANAL	18.2	14
GRAN CANAL	18.4	15
ENP 3	18.5	16
GRAN CANAL	18.6	17
ENP 3	18.9	18
GRAN CANAL	19.1	19
GRAN CANAL	19.3	20.5
ENP 3	19.3	20.5
ENP 3	21.4	22
ENP 3	22.0	23
ENP 3	22.9	24

3) SUMA DE RANGOS:

Para la estación Gran Canal:

$$R_1: 1 + 2 + 4 + 6 + 9.5 + 9.5 + 12.5 + 14 + 15 + 17 + 19 + 20.5 = 130$$

ó

En el caso de la ENP 3 sería

$$R_1: 3 + 5 + 7 + 8 + 11 + 12.5 + 16 + 18 + 20.5 + 22 + 23 + 24 = 170$$

4) Utilizando este último valor de R_1 , el valor de la estadística U de Mann-Whitney es

$$U = (12)(12) + \frac{(12)(13)}{2} - 170 = 52$$

5) Dado que $E(U) = (12)(12)/2 = 72$ y $Var(U) = (12)(12)(25)/12 = 300$, mediante la aproximación normal

$$Z = \frac{(52 - 72)}{\sqrt{300}}$$

$$Z = -1.15470$$

es el valor de una variable aleatoria normal estándar. Para $\alpha = 0.05$, los valores críticos son ± 1.96 . Por lo tanto *no puede rechazarse la hipótesis nula* de que las distribuciones provienen de distribuciones idénticas.

Con el programa *Statistica*, se puede realizar esta prueba no paramétrica muy fácilmente, se elaboró un macro (Ctrl + U) con los datos que se han manejado en este ejemplo para que sirva de referencia en la realización de otros casos de estudio, es importante mencionar que es requisito que el archivo tenga un formato como el que se muestra enseguida

	VAR1	VAR2
1	g-canal	13.5
2	g-canal	14.9
...		
12	g-canal	13.8
13	p3	14.3
...		
24	p3	15.2

Y por último es recomendable interpretar correctamente los datos de salida que genera el programa, recordando que el valor z es determinante en el rechazo de la prueba de hipótesis, mientras que la información que le acompaña son datos adicionales que se consideran en la prueba como se aprecia en el cuadro siguiente, sin embargo, no indica explícitamente si la prueba debe ser rechazada o no, esta decisión corresponde a la persona que realiza la prueba de hipótesis, teniendo siempre en mente los niveles de significancia y los criterios de rechazo para la prueba en particular.

	Suma de rangos G-Canal	Suma de rangos P3	U	Z	Nivel-p	Z Ajustada	p-level	N Gran canal	N p3
VAR2	130.0	170.0	52.0	-1.154	.2482	-1.15	.2479	12	12

Se confirmó lo ya observado en el ejemplo 2. Recordemos que en casos donde es factible utilizar la distribución normal, es más adecuado realizar pruebas paramétricas, aquí solo se mostró como una alternativa.

**RESULTADOS
Y
CONCLUSIONES**

.....

Para vivir en el mundo
se debe de estar despierto
jamás cerrarse los ojos
y desconfiar como un tuerto
pues todo es maldad en él,
son asechanzas y mañas
y por esto el que se duerme
amanece con lagañas

"Libro para el pueblo. Mil diez proverbios en verso para un mexicano"
Francisco Suarez Davila

RESULTADOS

EVALUACIÓN DE LOS OBJETIVOS DEL PROGRAMA

Se cubrieron los aspectos planteados como objetivo general pues los datos del clima recopilados en la estación meteorológica de la preparatoria número 3, en el período 1997-2001 fueron validados y organizados de manera que obtuvimos las estadísticas descriptivas de los principales elementos climatológicos; este proceso permitió a) detectar datos que proponen fallas en el funcionamiento del pluviómetro, b) verificar que los conceptos teóricos de la climatología son cumplidos por los datos correspondientes a la muestra, así como c) presentar la propuesta de elaborar publicaciones que surgen de dichas estadísticas. A lo largo de este proceso se dispuso el material de forma que fuera sencilla y de utilidad para el lector ya sea un estudiante o un profesor, orientando el interés hacia el conocimiento de los conceptos básicos de la estadística aplicada y en esta forma lograr nuestro objetivo específico, el logro de este último tendrá que ser evaluado cabalmente hasta que el material llegue a manos de los estudiantes del bachillerato, pues por el momento ha sido de utilidad únicamente para la coordinadora de la estación.

También revisamos la evolución histórica de software estadístico, así como los elementos que conforman los actuales programas de propósito general en esta materia, resaltando aquellas características que son determinantes en su elección. Utilizando *Statística* como herramienta en este trabajo se planteó su uso de forma muy concreta y pretendiendo ser didáctica hacia el conocimiento de aspectos del clima empleando la metodología estadística.

Finalmente se concretó un pequeño manual de ayuda al coordinador de estación, para realizar diferentes procesos que son de utilidad en su labor cotidiana, específicamente orientado a las actividades propuestas en el capítulo 5 y que representan algunas alternativas de análisis que pone a su disposición la estadística, tanto descriptiva como inferencial.

BENEFICIO A LA COMUNIDAD (APORTACIÓN A LA PROBLEMÁTICA MANEJADA)

Considerando que el análisis de los datos climatológicos en esta estación no había comenzado formalmente, este trabajo aportó algunas alternativas de análisis, tanto para el alumno como para el coordinador de la estación. Se investigaron datos y fuentes documentales de apoyo para la estación y que sirven también para su propia documentación; así como datos y procesos de validación y una propuesta para elaborar publicaciones como las que sugiere la OMM.

Tal vez la propuesta más importante y que puede retomarse, es la de utilizar la información producida con los datos, como material didáctico para la materia de geografía (para lograr un aprendizaje significativo, mediante la comprobación práctica

de los conocimientos teóricos) y principalmente como un medio para acercar a los estudiantes hacia las matemáticas aplicadas (específicamente a la estadística), sobre todo a aquellos que sientan inclinación por esta ciencia y al trabajo multidisciplinario.

FORMACIÓN PROFESIONAL

A partir de que tuve la oportunidad de colaborar en la Dirección General de Servicios de Cómputo Académico y mediante la realización de esta práctica profesional, pude retomar el área matemática, (que por cuestiones laborales no había podido ejercer), y ampliar o actualizar conocimientos referentes al campo del cómputo.

El re-aprender y aplicar la estadística representó un reto, pues después de varios años de "divorcio" nos habíamos alejado demasiado, aún cuando laboralmente siempre tuve inquietud por ponerla en práctica.

También el reconocer y trabajar el proceso de la Investigación y documentación científica, principalmente en cuanto a la labor de análisis y síntesis, representó un desafío profesional que trajo consigo un aprendizaje y toma de consciencia, sobre la importancia que tiene la metodología al investigar y hacer aportaciones para cualquier ciencia. Esto representó una motivación que fue clave durante la realización del trabajo.

Este trabajo influyó de manera muy importante en mi interés por la docencia de las matemáticas y especialmente de la estadística.

AUTOEVALUACIÓN

Lo único que puedo comentar (ya que cualquier otra cosa sería parcial), es que mi compromiso con este trabajo fue real.

OBSERVACIONES

MEJORAS AL PLAN DE ESTUDIOS

Tengo entendido que el plan de estudios de la licenciatura en Matemáticas Aplicadas y Computación ha cambiado recientemente, por lo cual no tengo elementos para hacer propuestas concretas, excepto que desearía que se incorporara un taller o laboratorio para la enseñanza del software de diversas materias como optimización y estadística, en las que existe un gran desarrollo de programas sobre el tema, tanto los creadores de dichos programas como los futuros profesionistas de las carreras de Matemáticas Aplicadas y Computación como de Actuaría (e inclusive de cualquier otra carrera que haga uso de la metodología estadística) se verían beneficiados si se realizan convenios con empresas o con instituciones educativas que han realizado dichos programas (pues de este modo darán a conocer sus productos a un sector más amplio).

Específicamente para la materia de Estadística, si el profesor pudiera diseñar, incorporar y, ejecutar prácticas de manera simultánea a sus alumnos, utilizando un software estadístico y una base de datos (sobre una red), se enriquecería el proceso de enseñanza - aprendizaje.

CONCLUSIONES

El estudio del clima es muy interesante pero nos ha proporcionado un pretexto excelente para estudiar algunos conceptos estadísticos y la posibilidad de su aplicación a dicha ciencia, lo cual hechos aprovechado en este informe.

La combinación clima-estadística puede tomar muchas direcciones de acuerdo a las necesidades de quien las estudie. Hemos ubicado al lector en el contexto de los esfuerzos realizados en cuanto a la recolección y análisis de datos climatológicos y pudimos ver que los archivos de este tipo pueden ser recopilados, validados, organizados y presentados para diferentes fines, algunas instancias como el Servicio Meteorológico Nacional y la Comisión Federal de Electricidad los utilizan para hacerlos del conocimiento público o como la materia prima para realizar pronósticos del tiempo, ayudados claro esta por software especializado, imágenes de satélite e infrarrojos, etc.; otras como la RAMA y el Instituto Mexicano del Petróleo para estudiar su relación con el comportamiento de los contaminantes en la zona metropolitana de la Ciudad de México*. O como en el caso del Centro de Ciencias de la Atmósfera (UNAM) en donde se realizan investigaciones para conocer el impacto del clima en los seres humanos y su entorno o verificar cambios climáticos.

En el presente trabajo los esfuerzos se dirigieron hacia un punto que no ha sido muy explotado, y es utilizar los datos muestrales del clima en la ENP 3, como un medio para interesar y acercar a estudiantes y coordinadores de estación (aún sin conocimientos formales) a los métodos de la estadística y al uso de alguna herramienta de software estadístico mostrando la riqueza de posibilidades de modo ameno y sencillo. Tal vez se logre un aprendizaje incidental o por lo menos se espera despertar el interés por la estadística básica aplicada.

También vimos como al paso del tiempo la utilización del cómputo ha impactado a muchas ciencias, por fortuna la estadística se cuenta entre ellas y a tenido grandes avances gracias a la simplificación de las tareas y a la reducción de tiempos en los cálculos de muchos métodos y procedimientos que el software estadístico trajo consigo. Esto se vio desde sus inicios con los primeros programas en plataforma mainframe como el BMD, SAS, SPSS y MINITAB, y hasta principios del siglo XXI con programas cada vez más amigables, de mayor potencia y rendimiento que están disponibles incluso gratuitamente por medio de Internet; pero gracias a esa disponibilidad que existe ahora nos enfrentamos a la dificultad de elegir entre tantas

* Prueba de esto son dos tesis de MAC realizadas con un enfoque de series de tiempo.

opciones, aquí revisamos algunas características que son consideradas como las más importantes en la elección de un software estadístico.

Investigamos con interés conceptos generales sobre el clima, los elementos que conforman la estación *Weather Monitor* que está instalada en las estaciones meteorológicas de las escuelas del bachillerato de la UNAM, y realizamos un estudio de la situación en la estación de la ENP 3. Completando la documentación de la estación con nuevos datos.

Respecto a los archivos brutos se estudió la mejor manera de exportarlos, pues existen necesidades de comparar o compartir datos entre estaciones o con otras instancias de la UNAM o externas, para esto se detallaron alternativas utilizando el programa *Statística* y haciendo uso de él, también se resolvió de manera muy fácil y rápida la dificultad que implicaba realizar la validación de los archivos climatológicos, a la que nunca se habían sometido los datos y que proponemos se realice regularmente.

A partir de ahí se explican de forma sencilla, tratando de ser claros y concretos, la metodología de organización y análisis que propone la estadística descriptiva, revisando conceptos como tablas de frecuencia y su representación geométrica mediante los histogramas y polígonos de frecuencia, ejemplificando con datos y el empleo del software estadístico. La necesidad de resumir las características de los elementos del clima nos llevó a reconocer los tipos de datos que se tienen en la muestra y estudiar las medidas numéricas descriptivas.

Planteamos desde un inicio que realizar un análisis estadístico descriptivo nos dirigía hacia la estadística inferencial, es decir, a estimar o realizar pruebas de hipótesis respecto a los parámetros de la población de la que surgió la muestra. Respecto a este punto, realizamos una breve descripción de los principales conceptos de la inferencia estadística tales como la relación entre distribuciones empíricas y teóricas (variables continuas) y su uso, el teorema del límite central, un ejemplo de estimación, así como la descripción de la distribución normal y la *t* de Student que nos sirvieron en los ejemplos de las pruebas de hipótesis como estadísticos de prueba. Finalizando con una opción de diferente para la prueba de dos medias, es decir, la prueba no paramétrica de Mann Whitney.

Apoyamos el trabajo con el desarrollo de macros que agilizan el trabajo en la estación, y además de solventar problemas de análisis que se tenían se propone aplicar los métodos estadísticos a otros problemas y campos de estudio.

Siempre intentamos provocar el interés del lector por la Estadística apoyados en su gusto por conocer más sobre el clima, tratando de mostrar cuán útil e interesante es la estadística aplicada.

Podemos concluir que debido a que la estadística es parte integral de la investigación científica así como fundamental en la toma de decisiones, el software estadístico tendrá que seguir sufriendo las transformaciones necesarias para adecuarse a los avances tecnológicos en cuanto a hardware y software se refiere.

REFERENCIAS

La inteligencia es una prostituta
que se vende por un poco de brillo
y que no sabe ya ruborizarse

Rosario Castellanos

REFERENCIAS

BIBLIOGRÁFICAS

Sobre estadística y software:

Canavos, George C.

- 1988 *Probabilidad y estadística. Aplicaciones y métodos*
1a. Edición en español (tr. Urbina Medal Edmundo Gerardo)
McGraw Hill. México.

Chou Ya-Lun

- 1972 *Análisis estadístico*
(tr. al español por Luis Eugenio di Marco)
Intercontinental. México.

Craddock J.M.

- 1968 *Statística in the computer age*
The English Universities Press Limited. London.

Davis, Gordon B. Y Olson, Margrethe H.

- 1987 *Sistemas de información gerencial*
Mc Graw-Hill. 2ª edición en español. México.

Elliot Alan C.

- 1985 "Assesing the impact of the microcomputer on statistics"
American Statistical Association. Proceedings of the Statistical Computing
Section. P. 168-169

HEPC

- 1993 "Computer and statistics in education. A brief history"
HEPC: Higher Education Product Companion.
Vol. 3, No. 2. Octubre. *Mathematics & quantitative tools*. P. 8-14.

Herrán, Nestor

- 1996 "7 programas estadísticos para Windows. Comparativa"
Byte España.
No.24, Época II. Diciembre. P 94-107.

Hoel, Paul G.

- 1988 *Estadística elemental*
6ª Impresión. Compañía Editorial Continental S.A. de C.V. México.

Huntsberger, David V. Y Billingsley Patrick

- 1983 *Elementos de estadística inferencial*
1a: Edición en español de la cuarta edición en Inglés (tr. Cervera Flores Armandina). Compañía Editorial Continental S.A. de C.V. México.

Koch, Armin y Haag, Uwe

- 1996 "The statistical software guide '95/96"
Computational statistics & data analysis
Vol. 21 No. 1. Febrero. P. 231-250
- 1997 "The statistical software guide '96/97"
Computational statistics & data analysis
Vol. 24. No. 1. Marzo. P. 107-131.

Kuonen Diego y Rohrl Armin S.A. (Interstat Online)

- 1988 "Identification of key steps to evaluate statistical software"
<http://interstat.stat.vt.edu/InterStat/ARTICLES/1998/abstracts/098002.html>
Octubre P. 1-21.

Lucas, Henry C.

1983 *Conceptos de los sistemas de información para la administración*
2ª. Ed (primera en español). Tr. Arturo García del Razo. Mc. Graw Hill.

Moore, Richard W.

1988 *Introduction to the use of computer packages for statistical analysis*
Prentice Hall Inc. Englewood New Jersey. P 3-4

Portus Govinden, Lincoyán

1984 *Curso práctico de estadística*
Mc Graw Hill. 1ª. Edición en español. México

Sanders, Donald H.

1990 *Informática: presente y futuro*
(tr. Roberto Luis Escalona)
Mc. Graw-Hill. 2ª edición en español. Colombia.

Statsoft Inc.

1993 *Statistica/w (Quick reference)*.
Tulsa, EUA.

Woodward Wayne A. & Elliot Alan C.

1982 "Statistical packages on microcomputers"
American Statistical Association. Proceedings of the Statistical Computing
Section. P. 63-70

Yacono, John J.

- 1994 "Anyone can be a statistician"
Windows Magazine
Noviembre. P 3 12

Sobre meteorología:

Brooks, Charles Ernest Pelham

- 1953 *Handbook of statistical methods in meteorology*
Her Majesty's stationery office. Londres.

Centro de Ciencias de la Atmósfera

- 1995 "Clima Urbano. Física de Nubes"
Ehecatl. Vol. 1. Núm. 1. México

ENP 3

- 1998 "El estado del tiempo en el país, durante los meses de marzo a abril"
Atmósfera 3. Año 5, No. 18. Marzo-mayo.

Escuela Nacional Preparatoria. Colegio de Ciencias y Humanidades. Programa de Integración de Docencia e Investigación. ENP-CCH-PIDI.

- 1995 Marco de referencia para la evaluación del programa de Estaciones Meteorológicas del Bachillerato Universitario

Mercado Mancera, Gustavo

- 1993 *Manual de prácticas : meteorología y climatología*
UNAM, FES Cuautitlán. México.

Organización Mundial Meteorológica (OMM) - SARH - SMN

- 1979 C. *Compendio de apuntes para la formación del personal meteorológico de la clase IV. Volumen 2. Climatología*
- 1979 M. *Compendio de apuntes para la formación del personal meteorológico de la clase IV. Volumen 3. Meteorología*
- 1980 *Compendio de apuntes para la formación del personal meteorológico de la clase III. Climatología*
- 1990 *Guía de prácticas climatológicas. OMM-Nº. 100*
(tr. de la 2ª ed. De la versión original inglesa publicada en 1983)

Red Automática de Monitoreo Atmosférico de la Ciudad de México (R.A.M.A.)

- 1995 Departamento del Distrito Federal.
Dirección General de Prevención y Control de la Contaminación
Red Automática de Monitoreo Atmosférico
México, diciembre

Del glosario y otros temas :

Carretero Mario
Constructivismo y Educación.
Ed. Edelvives.

Eco, Humberto

- 1994 *Cómo se hace una tesis. Técnicas y procedimientos de investigación, estudio y escritura*
1ª. edición en español de la 1ª. en italiano 1977 (tr. Baranda Lucía y Clavería Ibañez Alberto). Gedisa Editorial. España

REFERENCIAS

Díaz Barriga Arceo, Frida y Hernandez Rojas Gerardo

- 1996 *Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista.*
Ed. Mac Graw-Hill. México.

Freedman, Alan

- 1995 *The computer glossary*
7ª ed. American Management Association. USA.

Gooding Dan, Wang Wally y Van Buren Chris.

- 1995 *Diccionario ilustrado de computación para inexpertos*
Limusa S. A. de C.V.

Pfaffenberger Bryan

- 1995 *Diccionario para usuarios de computadoras*
5ª ed. (tr. Palmas Velasco Oscar)
Prentice-Hall Hispanoamericana S.A.

DE PÁGINAS DE INTERNET (abril 2002)

Sobre meteorología:

WWW, SMN Servicio Meteorológico Nacional

Dirección: <http://smn.cna.gob.mx/>

WWW, CCA Centro de Ciencias de la Atmósfera

Dirección: <http://serpiente.dgsc.unam.mx/cca/>

WWW, RAMA Red Automática de Monitoreo Atmosférico

Dirección: <http://www.sma.df.gob.mx/>

WWW, IMP Instituto Mexicano del Petróleo

Dirección: <http://www.imp.mx/>

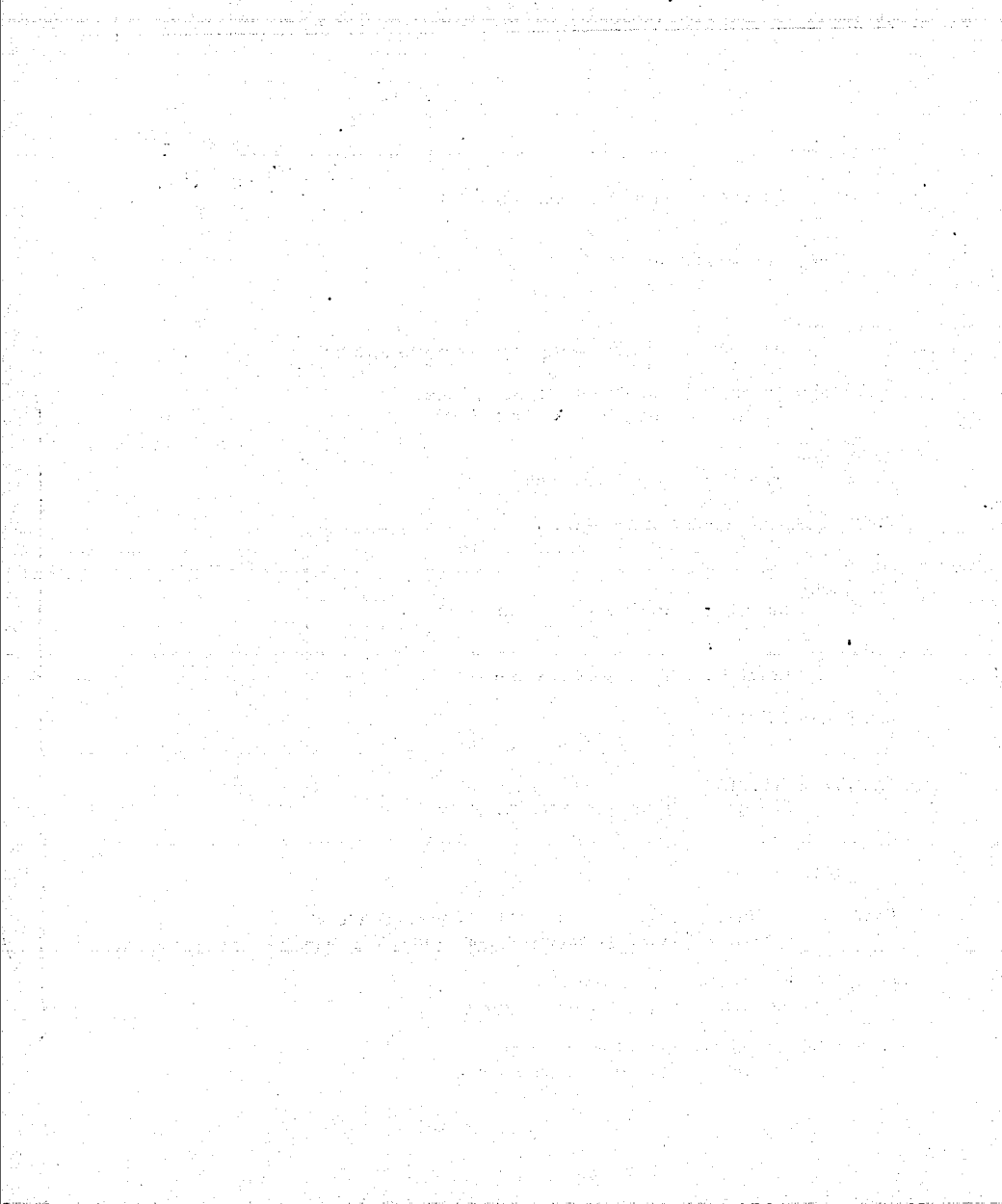
WWW, UNAMDirección: <http://www.unam.mx/>

Sobre software estadístico

WWW, GASPDirección: <http://www.stat.sc.edu/vllb/statstics.html>**WWW, ISI** International Association for Statistical ComputingDirección: <http://www.stat.unlpg.it/lasc/>**WWW, MINITAB**Dirección: <http://www.minitab.com/>**WWW, PHOENIX** (Distribuidor en México)Dirección: <http://www.mphoenix.com/>**WWW, SAS**Dirección: <http://www.sas.com/>**WWW, SPSS**Dirección: <http://www.spss.com/>**WWW, STATISTICA**Dirección: <http://www.statsoft.com/>**WWW, VENDEDORES**Dirección: <http://www.statstics.com/>

Otras:

WWW, DGSCA Dirección General de Servicios de Cómputo AcadémicoDirección: <http://www.museovirtual.unam.mx/dgsc.html>**WWW, ENP** Escuela Nacional PreparatoriaDirección: <http://dgenp.unam.mx>**WWW, CCH** Colegio de Ciencias y HumanidadesDirección: <http://coor.ccchs.unam.mx>



ANEXOS

A.	Barra de herramientas	A1
B.	Estadística descriptiva	B1
C.	Volumen de temperatura	C1

MANUAL

Importación	Ctrl + I	M1
Adecuar el archivo a importar	Ctrl + A	M3
Validación de datos	Ctrl + B	M9
Obtención de temperatura media diaria	Ctrl + E	M11
Obtención de estadística descriptiva	Ctrl + F	M13
Creación de tabla de contingencia	Ctrl + Z	M16
Gráficas de dispersión	Ctrl + G	M18
Comparación de dos medias	Ctrl + U	M20

Otros Procesos:

Gráficas en dos dimensiones	M22
Creación del archivo anual	M24
Elaboración de polígonos de frecuencia anuales	M25

GLOSARIO	G1
----------	----



Al estar de-seleccionado aparecen los números de caso y al seleccionarlo se ve el nombre que tiene asociado cada uno de ellos, nótese que en la columna más a la izquierda, la figura del lado izquierdo tiene un signo '#' y bajo él, el número de caso; mientras del lado derecho solo se ve (bajo la palabra 'Case') el nombre del renglón o caso.



Se incrementa el tamaño de la fuente y proporcionalmente todas las partes de la hoja de datos.




Decrece el tamaño de la fuente y las proporciones de la hoja de datos.

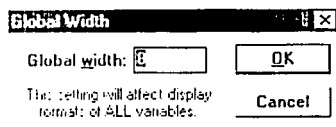








Incrementa el ancho de la columna resaltada en el momento

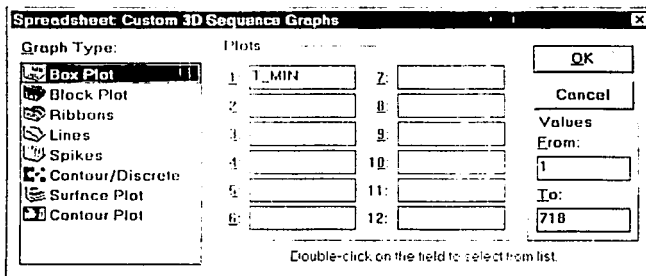


Decrece el ancho de la columna señalada en ese momento


-  Ajusta globalmente el ancho de las columnas mediante el cuadro de diálogo que sigue:

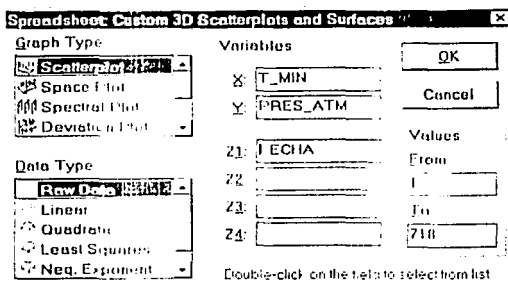



-  Aumenta una posición decimal al formato en que se despliega un valor numérico
-  Disminuye una posición decimal al formato en que se despliega un valor numérico
-  Recalcula un valor de una o varias variables para una fórmula definida previamente [ver Fig. M 8]
-  Aparece el manejador de valores texto (Text Value Manager)[ver Fig. M 9]
-  Crea distintas clases de gráficas en 2-D [ver Fig. M34]
-  Crea diferentes tipos de gráficas en 3-D, usando el cuadro de diálogo siguiente:




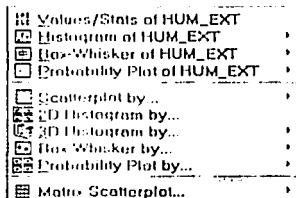
y especificando cualquier combinación de casos o variables, usa sus nombres como etiquetas para los ejes XY, visualizando finalmente una secuencia simple de datos.

 Genera gráficas de dispersión y “surfaces” (3-D). Aquí los valores representan coordenadas en un espacio tridimensional. Utiliza el siguiente cuadro de diálogo.

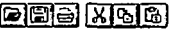


 Crea combinaciones de gráficas formando una matriz

 Despliega el menú de graficas estadísticas (Stats Graphs).



Difiere de las anteriores, ya que estas son herramientas de visualización propiamente estadísticas, como los histogramas.

 Su uso es idéntico que en Office; abrir, salvar e imprimir archivo y; cortar, copiar y pegar respectivamente, las partes seleccionadas de la hoja de datos.

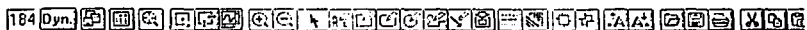
SCROLLSHEET (HOJA DE RESULTADOS)






-  Señala la celda actual con un color diferente al de las otras celdas, para resaltarla.
















Los otros botones funcionan igual que en la hoja de datos.

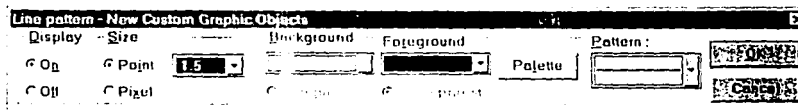
AMBIENTE GRÁFICO



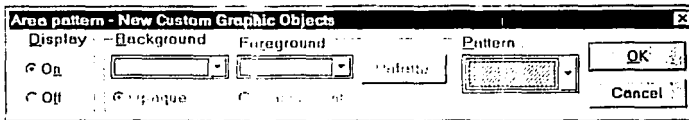
-  Indica las coordenadas XY en que se encuentra el cursor dentro de la ventana gráfica (modo dinámico)
-  Permite cambiar la escala de coordenadas a modo fijo (%), por ejemplo: la esquina inferior izquierda es la coordenada [0,0]% y la esquina superior derecha es la [100,100]%
-  Presenta el cuadro de diálogo para revisar o editar los datos en los que se basa la gráfica en pantalla. Ejemplo:

Temperature Media Data		
<input checked="" type="checkbox"/> Line Plot		
X	Y	
1.00	19.11	
2	18.90	
3	18.90	
4	19.20	
5	19.17	
6	19.40	
7	19.21	
8	19.14	
9	19.51	
10	19.47	
11	19.80	

-  Permite marcar o borrar temporalmente un punto de la gráfica, haciendo un clic sobre él
-  Este botón está activado por default y gracias a esto, al modificar el tamaño de una ventana gráfica, ésta mantiene sus proporciones.
-  Desactiva las funciones del botón anterior. Al estar éste activado y realizar cambios en el tamaño de la gráfica, sus elementos cambian de dimensiones
-  Al activar este botón aparece una barra de desplazamiento alrededor de la gráfica que habilita reposicionar la gráfica y el ajuste de sus márgenes.
-  Agranda el área seleccionada de la gráfica
-  Reduce el área seleccionada de la gráfica
-  Por default está activado y concede que se use el botón izquierdo del mouse para seleccionar un objeto o partes de la gráfica.
-  Acceso al modo texto para personalizar alguna leyenda dentro de la gráfica.
-  Dibuja un rectángulo.
-  Dibuja un rectángulo con los vértices redondeados
-  Crea un óvalo.
-  En este modo gráfico de manos libres, se pueden formar polígonos
-  Se dibujan diferentes tipos de flechas.
-  Activa el modo para ligar o insertar objetos gráficos desde un archivo
-  Selecciona el estilo de líneas, para el objeto seleccionado. Usa el cuadro de diálogo que sigue



- Selecciona el relleno para un objeto gráfico. Usa el cuadro de diálogo que aparece aquí:



- Cuando existen dos o más objetos encimados en la gráfica, al elegir uno de ellos y activar este botón, el objeto seleccionado queda al frente de todos.
- Cuando existen dos o más objetos encimados en la gráfica, al elegir uno de ellos y activar este botón, el objeto seleccionado queda atrás de todos.
- ▲ Incrementa proporcionalmente el tamaño de todos los tipos de letra en la gráfica
- ▲ Reduce proporcionalmente el tamaño de todos los tipos de letra en la gráfica

Otros botones para las gráficas en 3-dimensiones:

- Ajusta la perspectiva y rotación de la gráfica
- Despliega la gráfica por estratos, haciendo una pausa después de dibujar cada sección

TESIS CON
FALLA DE ORIGEN

ANEXO B

ESTADÍSTICAS DESCRIPTIVAS 1997

Enero

	n	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curstosis
T_INT	432	18.2	18.3	15.7	28.4	16.8	19.2	12.7	1.6	1.6	7.6
T_EXT	432	14.2	13.9	3.3	25.1	10.5	18.2	21.8	4.8	.0	-9
T_MAX	432	14.4	14.1	3.6	25.7	10.7	18.5	22.1	4.9	.1	-9
T_MIN	432	14.0	13.9	3.1	24.4	10.3	17.9	21.3	4.8	.0	-9
PRES_ATM	432	786.6	786.8	780.7	792.0	784.8	788.3	11.3	2.3	-1	-6
HUM_INT	432	31.2	33.0	12.0	45.0	25.0	37.0	33.0	8.5	-4	-8
HUM_EXT	432	41.6	38.5	7.0	86.0	27.0	57.0	79.0	18.7	.2	-9
P_RÓCIO	432	-2	1.8	-18.1	5.9	-3.1	4.1	24.0	5.5	-1.1	.1
VEL_VTO	432	1.5	.9	0.0	7.6	.4	2.2	7.6	1.5	1.6	2.4
RACHA	432	3.5	2.7	0.0	15.6	1.3	4.9	15.6	2.0	1.5	2.2
ENF_VTO	432	13.6	13.3	3.5	25.1	10.1	17.2	21.8	4.6	.2	-7
LLUVIA	432	.0	0.0	0.0	1.0	0.0	0.0	1.0	.0	20.8	432.0

Febrero

	n	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curstosis
T_INT	672	21.4	21.3	17.3	25.8	20.3	22.4	8.5	1.6	.2	-1
T_EXT	672	16.8	16.7	4.7	27.8	12.7	21.0	23.1	5.0	.1	-9
T_MAX	672	17.0	16.8	4.7	28.2	12.8	21.3	23.5	5.1	.1	-9
T_MIN	672	16.5	16.4	4.7	27.4	12.6	20.7	22.7	4.9	.1	-9
PRES_ATM	672	787.5	787.4	781.3	794.8	785.7	789.2	13.5	2.5	.2	-1
HUM_INT	672	32.4	33.0	18.0	44.0	30.0	35.0	26.0	4.8	-5	-4
HUM_EXT	672	45.0	44.0	11.0	86.0	32.0	58.0	75.0	16.4	.2	-8
P_RÓCIO	672	3.7	4.0	-8.1	8.8	2.1	5.6	16.9	2.7	-8	.9
VEL_VTO	672	1.4	1.3	0.0	5.8	.4	1.8	5.8	1.2	1.2	1.5
RACHA	672	3.4	2.7	0.0	10.7	1.8	4.5	10.7	2.2	1.0	.5
ENF_VTO	672	16.4	16.1	4.7	27.8	12.6	20.3	23.1	4.9	.1	-8
LLUVIA	672	.0	0.0	0.0	1.0	0.0	0.0	1.0	.0	25.9	672.0

Marzo

	n	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curstosis
T_INT	744	22.0	22.3	18.3	26.2	21.0	23.2	7.9	1.6	-3	-4
T_EXT	744	17.5	16.8	7.4	29.7	13.4	21.5	22.3	5.0	.3	-8
T_MAX	744	17.7	16.9	7.4	30.1	13.5	21.9	22.7	5.1	.3	-8
T_MIN	744	17.3	16.6	7.4	29.6	13.2	21.2	22.2	5.0	.3	-8
PRES_ATM	744	788.7	788.8	783.4	793.3	787.1	790.2	9.9	2.0	-2	-5
HUM_INT	744	34.6	36.0	14.0	49.0	31.0	40.0	35.0	6.9	-7	.0
HUM_EXT	744	48.6	49.0	10.0	94.0	32.0	65.5	84.0	19.8	.1	-1.1
P_RÓCIO	744	5.1	5.8	-6.4	10.8	3.5	7.6	17.2	3.4	-1.0	.6
VEL_VTO	744	1.6	1.3	0.0	7.6	.4	2.2	7.6	1.4	1.2	1.5
RACHA	744	3.7	3.1	0.0	13.0	1.8	4.9	13.0	2.5	1.1	.9
ENF_VTO	744	17.0	16.1	6.1	29.7	13.1	20.8	23.6	5.1	.4	-7
LLUVIA	744	.0	0.0	0.0	11.0	0.0	0.0	11.0	.5	19.7	424.2

**TESIS CON
FALLA DE ORIGEN**

Abril

	Nº de Observaciones	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	600	23.1	22.9	19.5	27.3	22.3	23.7	7.8	1.3	.9	.9
T_EXT	600	18.2	17.7	9.7	29.3	14.3	21.8	19.6	4.5	.3	-9
T_MAX	600	18.5	18.0	9.8	29.6	14.4	22.1	19.8	4.6	.3	-9
T_MIN	600	17.9	17.4	9.6	28.9	14.1	21.4	19.3	4.5	.3	-8
PRES_ATM	600	787.1	787.2	781.0	791.1	785.6	789.0	10.1	2.3	-.4	-.5
HUM_INT	600	34.0	36.0	13.0	49.0	27.0	41.0	36.0	8.2	-.5	-8
HUM_EXT	600	48.1	45.0	9.0	91.0	33.0	65.0	82.0	20.2	.2	-1.0
P_RÓCIO	600	5.5	7.1	-8.4	12.1	2.1	8.7	20.5	4.2	-1.0	.1
VEL_VTO	600	1.6	1.3	0.0	6.7	.4	2.2	6.7	1.4	1.1	1.2
RACHA	600	4.0	3.1	0.0	13.9	1.8	5.8	13.9	2.8	1.0	.4
ENF_VTO	600	17.7	17.2	7.6	29.1	14.1	20.6	21.5	4.4	.4	-.6
LLUVIA	600	.1	0.0	0.0	12.0	0.0	0.0	12.0	.7	12.3	166.3

Mayo

	Nº de Observaciones	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	743	24.8	24.6	21.7	28.7	23.8	25.4	7.0	1.3	.7	.3
T_EXT	743	17.9	17.2	10.7	30.0	14.2	21.5	19.3	4.3	.5	-.7
T_MAX	743	18.3	17.5	10.8	30.7	14.4	22.1	19.9	4.5	.5	-.8
T_MIN	743	17.5	16.7	10.4	29.3	14.1	20.9	18.9	4.2	.5	-.7
PRES_ATM	743	789.6	789.7	784.0	795.6	788.6	790.7	11.6	1.8	.1	1.0
HUM_INT	743	32.6	33.0	18.0	45.0	30.0	36.0	27.0	5.3	-.6	-.1
HUM_EXT	743	55.5	57.0	14.0	90.0	38.0	72.0	76.0	19.7	-.1	-1.1
P_RÓCIO	743	7.7	8.3	-1.6	11.7	6.5	9.8	13.3	2.7	-1.1	.6
VEL_VTO	743	1.6	1.3	0.0	7.6	.4	2.2	7.6	1.4	1.2	1.3
RACHA	743	4.2	3.6	0.0	13.9	2.2	5.8	13.9	2.7	.9	.4
ENF_VTO	743	17.3	16.4	8.8	30.0	13.8	20.7	21.2	4.3	.6	-.5
LLUVIA	743	.2	0.0	0.0	55.0	0.0	0.0	55.0	2.2	21.2	516.0

Junio

	Nº de Observaciones	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	717	26.8	26.5	23.8	30.8	25.4	27.9	7.0	1.7	.4	-.8
T_EXT	717	19.8	19.1	12.2	30.2	16.3	23.5	18.0	4.3	-.4	-1.0
T_MAX	717	20.3	19.4	12.2	30.7	16.5	24.1	18.5	4.5	-.4	-1.0
T_MIN	717	19.5	18.6	12.1	29.9	15.9	23.0	17.8	4.2	-.4	-.9
PRES_ATM	717	788.7	788.9	782.5	792.9	787.5	790.0	10.4	1.8	-.4	.0
HUM_INT	717	33.9	34.0	18.0	44.0	32.0	38.0	26.0	4.9	-.8	-.6
HUM_EXT	717	57.0	60.0	16.0	98.0	41.0	72.0	82.0	19.6	-.2	-.9
P_RÓCIO	717	10.0	10.9	.7	14.0	8.9	11.9	13.3	2.7	-1.4	1.6
VEL_VTO	717	1.8	1.3	0.0	6.7	.9	2.7	6.7	1.3	1.0	.5
RACHA	717	4.3	3.6	0.0	14.3	2.2	5.8	14.3	2.5	.8	.1
ENF_VTO	717	19.4	18.4	11.3	29.7	15.9	22.7	18.4	4.2	.5	-.9
LLUVIA	717	.2	0.0	0.0	41.0	0.0	0.0	41.0	1.9	16.1	309.4

Julio

	Tempo	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	743	24.6	24.5	22.8	26.9	23.9	25.3	4.1	.9	.3	-.8
T_EXT	743	18.3	17.1	12.4	27.4	15.0	21.6	15.0	3.9	.6	-.9
T_MAX	743	18.7	17.6	12.6	28.4	15.2	22.2	15.8	4.1	.5	-1.0
T_MIN	743	17.9	16.8	12.3	26.6	14.8	20.8	14.3	3.7	.6	-.8
PRES_ATM	743	791.1	791.3	786.6	794.0	790.1	792.2	7.4	1.5	-.5	-.4
HUM_INT	743	39.5	39.0	28.0	47.0	37.0	43.0	19.0	3.7	-.2	-.4
HUM_EXT	743	67.7	71.0	24.0	100.0	53.0	83.0	76.0	18.1	-.4	-.9
P_ROCIO	743	11.5	11.7	4.3	14.5	10.7	12.6	10.2	1.5	-1.1	2.4
VEL_VTO	743	1.6	1.3	0.0	7.2	.9	2.2	7.2	1.4	1.4	1.6
RACHA	743	4.1	3.6	.4	13.4	2.2	5.4	13.0	2.5	1.1	.9
ENF_VTO	743	17.8	16.4	8.3	27.4	14.8	20.6	19.1	3.7	.7	-.7
LLUVIA	743	.4	0.0	0.0	57.0	0.0	0.0	57.0	3.4	11.4	150.1

Agosto

	Tempo	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	743	24.7	24.7	21.9	27.1	23.9	25.4	5.2	1.1	.0	-.7
T_EXT	743	18.5	17.4	12.1	28.3	15.1	22.0	16.2	4.1	.5	-.9
T_MAX	743	18.9	17.9	12.2	28.7	15.2	22.6	16.5	4.3	.5	-1.0
T_MIN	743	18.1	17.1	11.6	27.9	14.9	21.3	16.3	3.9	.6	-.8
PRES_ATM	743	791.8	792.1	786.2	795.1	790.8	792.9	8.9	1.5	-.7	-.4
HUM_INT	743	37.6	38.0	24.0	48.0	35.0	40.0	24.0	4.1	-.3	.0
HUM_EXT	743	63.1	66.0	19.0	96.0	48.0	79.0	77.0	18.7	-.4	-.9
P_ROCIO	743	10.5	10.9	2.4	14.3	9.5	11.9	11.9	1.9	-1.2	1.6
VEL_VTO	743	1.8	1.3	0.0	6.7	.9	2.7	6.7	1.5	1.2	.9
RACHA	743	4.3	3.6	0.0	14.8	2.2	5.8	14.8	2.7	.9	.3
ENF_VTO	743	17.9	16.6	12.1	28.1	14.7	21.1	16.0	4.0	.7	-.7
LLUVIA	743	.2	0.0	0.0	24.0	0.0	0.0	24.0	1.4	13.0	195.8

Septiembre

	Tempo	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	709	22.3	22.2	20.3	25.9	21.7	22.8	5.6	.9	.5	.6
T_EXT	709	18.2	17.1	12.1	27.6	15.2	20.6	15.5	3.7	.7	-.6
T_MAX	709	18.6	17.4	12.2	29.1	15.4	21.3	16.9	3.9	.7	-.6
T_MIN	709	17.8	16.7	12.0	26.8	15.1	20.1	14.8	3.4	.7	-.5
PRES_ATM	709	790.6	790.7	785.5	794.6	789.7	791.6	9.1	1.6	-.3	.2
HUM_INT	709	47.7	48.0	33.0	61.0	44.0	51.0	28.0	5.4	-.2	-.4
HUM_EXT	709	66.9	71.0	28.0	100.0	55.0	78.5	72.0	16.4	-.4	-.6
P_ROCIO	709	11.4	11.3	5.3	15.3	10.5	12.6	10.0	1.6	-.7	1.3
VEL_VTO	709	1.6	1.3	0.0	6.3	.9	2.2	6.3	1.2	1.2	1.1
RACHA	709	3.9	3.6	0.0	13.0	2.2	4.9	13.0	2.2	1.0	.9
ENF_VTO	709	17.7	16.3	11.4	27.6	14.8	19.9	16.2	3.7	.9	-.3
LLUVIA	709	.2	0.0	0.0	57.0	0.0	0.0	57.0	2.4	19.8	447.5

Octubre

	#Observaciones	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	743	21.6	22.2	15.6	25.8	20.4	23.2	10.2	2.1	-7	-3
T_EXT	743	17.2	16.6	3.4	28.4	14.1	20.9	25.0	4.9	0	-5
T_MAX	743	17.5	16.8	3.8	29.2	14.3	21.6	25.4	5.1	1	-6
T_MIN	743	16.8	16.3	3.2	27.6	13.8	20.5	24.4	4.8	0	-5
PRES_ATM	743	790.3	790.4	783.6	795.8	788.7	792.2	12.2	2.3	-1	-7
HUM_INT	743	43.0	43.0	25.0	60.0	38.0	48.0	35.0	6.9	-3	-4
HUM_EXT	743	60.5	64.0	15.0	97.0	48.0	75.0	82.0	18.7	-5	-6
P_ROCIO	743	8.5	9.4	-5.7	15.2	6.0	11.7	20.9	4.0	-9	4
VEL_VTO	743	1.4	.9	0.0	6.3	.4	1.8	6.3	1.2	1.2	1.1
RACHA	743	3.5	3.1	0.0	11.2	1.8	4.9	11.2	2.2	.8	1
ENF_VTO	743	16.8	16.2	3.4	28.4	13.6	20.3	25.0	4.9	2	-4
LLUVIA	743	.1	0.0	0.0	8.0	0.0	0.0	8.0	.5	11.3	148.8

Noviembre

	#Observaciones	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	720	21.6	21.5	17.4	25.9	20.7	22.5	8.5	1.4	1	1
T_EXT	720	16.6	16.2	7.9	27.2	13.3	19.9	19.3	4.3	2	-8
T_MAX	720	16.9	16.6	8.2	27.7	13.6	20.4	19.5	4.4	2	-8
T_MIN	720	16.2	15.8	7.7	26.6	13.1	19.4	18.9	4.2	2	-7
PRES_ATM	720	791.5	791.5	785.1	797.1	789.6	793.6	12.0	2.4	-0	-8
HUM_INT	720	38.7	39.0	19.0	55.0	34.0	44.0	36.0	6.4	-3	-2
HUM_EXT	720	57.4	58.5	13.0	90.0	45.0	72.0	77.0	17.7	-3	-7
P_ROCIO	720	7.2	7.8	-5.2	13.6	5.8	9.2	18.8	2.9	-1.0	1.4
VEL_VTO	720	1.2	.9	0.0	5.8	.4	1.8	5.8	1.1	1.3	1.8
RACHA	720	3.2	2.7	0.0	10.7	1.8	4.0	10.7	2.0	1.0	7
ENF_VTO	720	16.3	15.8	7.9	27.2	13.2	19.1	19.3	4.2	3	-6
LLUVIA	720	.0	0.0	0.0	12.0	0.0	0.0	12.0	.5	25.4	663.9

Diciembre

	#Observaciones	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis
T_INT	725	19.2	18.7	14.2	24.7	17.4	21.3	10.5	2.3	2	-9
T_EXT	725	15.8	15.3	1.4	28.0	12.1	19.3	26.6	5.4	1	-3
T_MAX	725	16.2	15.6	1.5	28.5	12.4	19.9	27.0	5.5	1	-3
T_MIN	725	15.4	14.9	1.3	27.6	11.8	18.8	26.3	5.3	0	-2
PRES_ATM	725	791.4	791.5	784.6	798.9	789.9	792.9	14.3	2.5	0	-1
HUM_INT	725	36.5	36.0	17.0	55.0	31.0	42.0	38.0	8.4	-0	-6
HUM_EXT	725	50.5	53.0	10.0	94.0	35.0	65.0	84.0	19.1	-2	-8
P_ROCIO	725	4.2	5.9	-11.1	12.0	1.4	7.7	23.1	4.8	-1.0	3
VEL_VTO	725	1.4	.9	0.0	8.0	.4	1.8	8.0	1.4	1.7	3.3
RACHA	725	3.6	2.7	0.0	15.2	1.8	4.5	15.2	2.7	1.5	1.9
ENF_VTO	725	15.3	14.7	1.4	27.9	11.6	18.8	26.5	5.5	1	-3
LLUVIA	725	.0	0.0	0.0	2.0	0.0	0.0	2.0	.1	12.0	155.1

ESTADÍSTICAS DESCRIPTIVAS 1998

Enero

	N	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curstosis
T_INT	744	18.1	18.2	14.4	21.6	17.2	19.1	7.2	1.5	-.3	-.3
T_EXT	744	14.3	13.8	4.8	26.6	10.2	18.6	21.8	5.2	.3	-9
T_MAX	744	14.7	14.2	4.9	27.3	10.4	19.2	22.4	5.3	.3	-1.0
T_MIN	744	13.9	13.4	4.6	26.0	9.8	17.9	21.4	5.2	.3	-9
PRES_ATM	744	792.6	792.7	786.3	798.9	790.9	794.1	12.6	2.4	.0	-.3
HUM_INT	744	31.5	32.0	15.0	50.0	27.0	36.0	35.0	6.4	.0	-1
HUM_EXT	744	44.7	43.0	10.0	91.0	28.0	59.0	81.0	19.5	.3	-8
P_ROCIO	744	.9	1.2	-12.0	9.0	-.9	3.4	21.0	3.5	-.6	.7
VEL_VTO	744	1.2	.9	0.0	5.8	.4	1.8	5.8	1.1	1.2	1.3
RACHA	744	3.3	2.7	0.0	12.1	1.8	4.5	12.1	2.2	1.0	.6
ENF_VTO	744	14.0	13.6	4.8	26.6	10.1	17.8	21.8	5.1	.3	-9
LLUVIA	744	.0	0.0	0.0	2.0	0.0	0.0	2.0	.1	13.5	186.1

Febrero

	N	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curstosis
T_INT	554	19.4	19.2	15.1	24.2	18.1	20.8	9.1	1.8	.2	-.7
T_EXT	554	15.7	15.1	2.8	27.5	10.9	20.6	24.7	5.9	.2	-1.0
T_MAX	554	16.3	15.6	3.1	28.4	11.5	21.2	25.3	6.0	-.1	-1.0
T_MIN	554	15.2	14.7	2.1	27.0	10.5	19.9	24.9	5.8	.2	-.9
PRES_ATM	554	792.1	792.5	785.0	799.8	790.6	793.8	14.8	2.6	-.4	-1
HUM_INT	554	21.3	21.0	9.0	36.0	17.0	26.0	27.0	6.0	.0	-.7
HUM_EXT	554	29.3	27.0	7.0	73.0	17.0	39.0	66.0	15.0	.4	-.6
P_ROCIO	554	-4.1	-4.3	-14.0	4.9	-7.2	-.6	18.9	4.4	-.0	-.8
VEL_VTO	554	1.8	1.3	0.0	7.2	.4	2.7	7.2	1.5	1.0	.1
RACHA	554	4.7	3.6	0.0	17.4	1.8	7.2	17.4	3.4	1.0	-.4
ENF_VTO	554	15.0	14.2	2.8	27.5	10.5	19.3	24.7	5.6	.2	-.9
LLUVIA	554	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		

Marzo

	N	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curstosis
T_INT	741	22.5	22.4	17.0	27.4	21.0	23.9	10.4	1.9	.0	-.5
T_EXT	741	18.9	18.8	5.3	30.9	14.2	24.0	25.6	5.8	-.1	-1.0
T_MAX	741	19.7	19.7	6.5	31.6	14.8	24.8	25.1	5.9	-.1	-1.0
T_MIN	741	18.1	18.0	4.7	30.6	13.3	23.1	25.9	5.8	-.0	-1.0
PRES_ATM	741	795.0	794.7	788.8	800.7	793.2	797.0	11.9	2.6	.2	-.7
HUM_INT	741	22.8	23.0	8.0	39.0	18.0	27.0	31.0	6.4	.2	-.4
HUM_EXT	741	30.0	27.0	7.0	79.0	19.0	39.0	72.0	15.5	1.0	.6
P_ROCIO	741	-.8	-.1	-14.7	8.3	-3.8	2.6	23.0	4.5	-.6	-.1
VEL_VTO	741	1.7	1.3	0.0	6.3	.4	2.7	6.3	1.3	.8	-.2
RACHA	741	4.8	4.0	0.0	13.0	2.2	7.2	13.0	3.0	.7	-.5
ENF_VTO	741	18.5	18.2	5.3	30.7	14.1	23.2	25.4	5.6	.0	-.9
LLUVIA	741	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		

Abril

	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis	
T_INT	715	26.0	26.2	21.1	30.0	24.7	27.2	8.9	1.8	-3	-6
T_EXT	715	22.0	21.8	11.0	32.7	17.6	26.3	21.7	5.4	.1	-1.0
T_MAX	715	22.5	22.3	11.2	33.3	18.1	27.1	22.1	5.4	.0	-1.0
T_MIN	715	21.4	21.2	10.9	32.3	17.0	25.6	21.4	5.3	.1	-1.0
PRES_ATM	715	796.7	796.6	792.0	802.0	795.4	797.9	10.0	2.0	.1	-.2
HUM_INT	715	26.2	26.0	11.0	46.0	22.0	30.0	35.0	6.1	.3	.2
HUM_EXT	715	33.4	30.0	9.0	81.0	23.0	42.0	72.0	15.7	1.0	.5
P.ROCIO	715	3.6	4.1	-7.2	10.7	1.4	6.1	17.9	3.4	-.5	-.0
VEL_VTO	715	1.5	1.3	0.0	6.7	.4	2.2	6.7	1.3	1.1	.8
RACHA	715	4.2	3.6	0.0	13.0	1.8	6.3	13.0	2.7	.8	-.2
ENF_VTO	715	21.7	21.4	11.0	32.7	17.6	25.8	21.7	5.3	.1	-1.0
LLUVIA	715	.0	0.0	0.0	18.0	0.0	0.0	18.0	.7	25.1	653.6

Mayo

	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis	
T_INT	744	26.4	26.5	22.9	29.7	25.3	27.5	6.8	1.4	-1	-6
T_EXT	744	22.9	22.5	11.8	34.0	18.3	27.7	22.2	5.4	.1	-1.1
T_MAX	744	23.7	23.4	12.3	34.8	19.1	28.8	22.5	5.5	.1	-1.2
T_MIN	744	22.1	21.7	11.0	33.7	17.7	26.8	22.7	5.3	.2	-1.1
PRES_ATM	744	798.6	798.5	793.6	803.3	797.3	800.1	9.9	2.0	.0	-.4
HUM_INT	744	24.1	24.0	13.0	43.0	20.0	27.0	30.0	5.0	.5	.2
HUM_EXT	744	29.7	27.0	9.0	71.0	19.0	38.0	62.0	13.6	.7	-.1
P.ROCIO	744	2.8	2.5	-5.6	11.9	.2	5.6	17.5	3.6	.1	-.5
VEL_VTO	744	1.6	1.3	0.0	7.6	.4	2.2	7.6	1.5	1.4	1.7
RACHA	744	4.6	3.6	0.0	16.1	2.2	6.7	16.1	3.0	.9	.3
ENF_VTO	744	22.6	22.1	11.8	34.0	18.2	27.2	22.2	5.4	.2	-1.1
LLUVIA	744	.0	0.0	0.0	5.0	0.0	0.0	5.0	.2	22.2	529.6

Junio

	Media	Mediana	Mínimo	Máximo	Q1	Q3	Rango	Des-est.	Simetría	Curiosis	
T_INT	720	25.7	25.8	22.7	29.4	24.8	26.7	6.7	1.3	-3	-6
T_EXT	720	21.4	20.6	13.9	32.6	17.5	25.1	18.7	4.6	-.4	-1.0
T_MAX	720	22.1	21.4	14.1	33.6	17.9	26.3	19.5	4.8	.3	-1.1
T_MIN	720	20.7	19.7	13.4	31.7	17.1	24.1	18.3	4.4	.5	-.8
PRES_ATM	720	798.7	798.9	793.6	802.3	797.6	799.7	8.7	1.6	-.3	-.1
HUM_INT	720	39.0	39.0	17.0	55.0	34.0	45.0	38.0	7.0	-.0	-.2
HUM_EXT	720	54.6	55.0	10.0	96.0	38.0	71.0	86.0	20.1	-.0	-.9
P.ROCIO	720	10.7	11.4	-3.2	15.4	9.2	12.8	18.6	3.0	-1.4	2.7
VEL_VTO	720	1.7	1.3	0.0	6.3	.9	2.2	6.3	1.3	1.3	1.4
RACHA	720	4.9	4.5	.4	18.8	2.7	6.3	18.4	2.7	1.1	1.3
ENF_VTO	720	21.0	20.2	13.1	32.6	17.3	24.4	19.5	4.5	.5	-.8
LLUVIA	720	.3	0.0	0.0	15.0	0.0	0.0	15.0	1.6	6.8	49.4

Julio

	Nº de	Media:	Mediana:	Mínimo	Máximo	Q1:	Q3:	Rango:	Des-est:	Simetría	Curstosis
T_INT	744	23.1	23.1	20.1	25.7	22.3	23.8	5.6	1.0	.0	-.3
T_EXT	744	19.3	18.3	12.6	29.4	16.2	22.3	16.8	3.9	.5	-.8
T_MAX	744	20.1	19.1	12.7	30.3	16.6	23.4	17.6	4.2	.5	-.9
T_MIN	744	18.6	17.7	12.2	28.4	15.8	21.1	16.2	3.6	.6	-.6
PRES_ATM	744	798.2	798.4	794.0	801.1	797.2	799.2	7.1	1.4	-.4	-.4
HUM_INT	744	44.7	45.5	31.0	55.0	41.0	49.0	24.0	5.4	-.4	-.6
HUM_EXT	744	61.3	65.0	21.0	96.0	48.5	75.0	75.0	17.3	-.4	-.8
P_ROCIO	744	10.9	11.5	4.0	14.7	10.1	12.4	10.7	2.1	-.8	.4
VEL_VTO	744	1.9	1.3	0.0	7.2	.9	2.7	7.2	1.4	1.3	1.6
RACHA	744	5.1	4.5	.4	16.1	3.1	6.7	15.7	2.8	1.0	.8
ENF_VTO	744	18.7	17.7	10.6	29.4	15.7	21.7	18.8	3.9	.6	-.6
LLUVIA	744	.3	0.0	0.0	25.0	0.0	0.0	25.0	1.9	9.1	95.6

Agosto

	Nº de	Media:	Mediana:	Mínimo	Máximo	Q1:	Q3:	Rango:	Des-est:	Simetría	Curstosis
T_INT	739	22.6	22.5	20.8	25.8	21.9	23.2	5.0	1.0	.7	-.4
T_EXT	739	18.5	17.5	11.8	27.6	15.6	21.4	15.8	3.6	.5	-.8
T_MAX	739	19.2	18.1	11.9	28.2	15.8	22.7	16.3	4.0	.5	-1.0
T_MIN	739	17.8	16.8	11.3	26.3	15.2	20.3	15.0	3.3	.6	-.7
PRES_ATM	739	797.4	797.5	792.0	801.6	795.8	799.0	9.6	2.1	-.2	-.6
HUM_INT	739	49.3	49.0	38.0	65.0	46.0	53.0	27.0	4.5	-.1	-.4
HUM_EXT	739	68.4	72.0	33.0	97.0	55.0	82.0	64.0	16.7	-.3	-1.0
P_ROCIO	739	12.0	12.2	7.3	15.2	11.2	13.0	7.9	1.4	-.4	-.2
VEL_VTO	739	1.6	1.3	0.0	5.8	.9	2.2	5.8	1.3	1.1	.6
RACHA	739	4.5	4.0	.4	15.6	2.7	5.8	15.2	2.5	1.0	.6
ENF_VTO	739	18.0	16.8	11.8	27.6	15.2	20.8	15.8	3.6	.7	-.6
LLUVIA	739	.7	0.0	0.0	50.0	0.0	0.0	50.0	4.0	8.5	85.7

Septiembre

	Nº de	Media:	Mediana:	Mínimo	Máximo	Q1:	Q3:	Rango:	Des-est:	Simetría	Curstosis
T_INT	718	22.7	22.5	20.3	25.8	21.8	23.4	5.5	1.1	.5	-.5
T_EXT	718	18.1	17.0	13.6	25.6	15.7	20.6	12.0	3.0	.7	-.7
T_MAX	718	18.7	17.4	13.8	26.7	15.9	21.6	12.9	3.4	.7	-.8
T_MIN	718	17.6	16.7	13.3	24.9	15.5	19.5	11.6	2.7	.8	-.5
PRES_ATM	718	792.9	793.0	788.8	796.1	791.7	794.2	7.3	1.6	-.3	-.6
HUM_INT	718	54.9	56.0	36.0	69.0	50.0	60.0	33.0	6.4	-.2	-.7
HUM_EXT	718	77.6	81.0	38.0	100.0	64.0	90.0	62.0	15.4	-.4	-1.0
P_ROCIO	718	13.8	13.9	9.1	16.7	13.2	14.6	7.6	1.1	-.5	.5
VEL_VTO	718	1.4	.9	0.0	5.4	.4	1.8	5.4	1.1	1.1	.7
RACHA	718	3.9	3.6	.4	17.9	2.2	4.9	17.5	2.3	1.3	2.8
ENF_VTO	718	17.7	16.6	12.3	25.6	15.6	20.1	13.3	2.9	.8	-.4
LLUVIA	718	1.2	0.0	0.0	53.0	0.0	0.0	53.0	5.1	7.0	56.5

Octubre

	742	20.7	20.9	17.6	23.8	20.0	21.6	6.2	1.3	-5	-2
T_INT	742	20.7	20.9	17.6	23.8	20.0	21.6	6.2	1.3	-5	-2
T_EXT	742	16.6	15.8	8.6	25.4	14.1	18.9	16.8	3.6	.6	-3
T_MAX	742	17.3	16.2	8.9	27.3	14.3	19.9	18.4	3.9	.6	-5
T_MIN	742	16.1	15.3	7.8	24.8	13.8	18.1	17.0	3.3	.6	-1
PRES_ATM	742	794.5	794.4	789.2	799.8	793.1	795.8	10.6	2.0	.1	-2
HUM_INT	742	55.4	56.0	30.0	68.0	52.0	59.0	38.0	5.6	-1.0	2.5
HUM_EXT	742	75.7	78.0	19.0	100.0	66.0	88.0	81.0	16.1	-8	.2
P. ROCÍO	742	11.9	12.2	.4	16.6	10.7	13.3	16.2	2.1	-1.0	2.6
VEL_VTO	742	1.4	1.3	0.0	5.8	.4	2.2	5.8	1.1	.9	.6
RACHA	742	4.0	3.6	.4	11.2	2.2	5.4	10.8	2.1	.6	-0
ENF_VTO	742	16.2	15.2	8.6	25.4	13.7	18.2	16.8	3.6	.7	-1
LLUVIA	742	.4	0.0	0.0	44.0	0.0	0.0	44.0	2.8	9.8	116.3

Noviembre

	720	21.2	21.2	17.9	24.2	20.4	22.0	6.3	1.2	-0	-1
T_INT	720	21.2	21.2	17.9	24.2	20.4	22.0	6.3	1.2	-0	-1
T_EXT	720	17.4	16.6	7.1	27.6	13.6	21.7	20.5	4.6	.2	-1.1
T_MAX	720	18.2	17.3	7.2	29.4	14.1	22.9	22.2	4.9	.2	-1.1
T_MIN	720	16.7	15.9	6.8	26.6	13.2	20.7	19.8	4.4	.2	-1.0
PRES_ATM	720	794.6	795.0	787.3	799.0	792.8	796.4	11.7	2.5	-5	-4
HUM_INT	720	45.0	46.0	31.0	59.0	41.0	49.0	28.0	5.5	-4	-6
HUM_EXT	720	62.1	65.0	23.0	100.0	45.0	79.0	77.0	19.7	-2	-1.1
P. ROCÍO	720	9.2	9.8	-.4	14.9	7.4	11.1	15.3	2.6	-7	.2
VEL_VTO	720	1.3	.9	0.0	5.4	.4	1.8	5.4	1.1	1.4	1.7
RACHA	720	3.7	3.1	.4	13.4	2.2	4.5	13.0	2.1	1.1	1.1
ENF_VTO	720	17.1	16.0	7.1	27.6	13.4	21.3	20.5	4.6	.3	-1.0
LLUVIA	720	.1	0.0	0.0	39.0	0.0	0.0	39.0	1.9	17.0	301.7

Diciembre

	744	18.2	17.8	15.5	21.9	16.9	19.4	6.4	1.5	.4	-9
T_INT	744	18.2	17.8	15.5	21.9	16.9	19.4	6.4	1.5	.4	-9
T_EXT	744	15.2	14.4	4.5	25.8	10.2	20.5	21.3	5.6	.1	-1.3
T_MAX	744	16.1	15.4	4.8	28.1	10.9	21.8	23.3	5.9	.1	-1.3
T_MIN	744	14.3	13.5	4.1	25.6	9.4	19.5	21.5	5.5	.1	-1.3
PRES_ATM	744	796.9	797.0	791.7	801.3	795.6	798.5	9.6	2.0	-2	-5
HUM_INT	744	39.2	40.0	25.0	49.0	36.0	42.0	24.0	4.6	-2	-6
HUM_EXT	744	54.4	55.0	18.0	94.0	36.0	72.0	76.0	20.1	-0	-1.3
P. ROCÍO	744	4.8	5.1	-4.1	12.0	3.0	6.8	16.1	2.8	-4	-1
VEL_VTO	744	-.9	.9	0.0	4.0	.4	1.3	4.0	.8	1.3	1.9
RACHA	744	2.8	2.7	0.0	8.9	1.8	3.6	8.9	1.7	1.0	.9
ENF_VTO	744	15.1	14.3	4.5	25.8	10.2	20.3	21.3	5.6	.1	-1.2
LLUVIA	744	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		

ANEXO C

VOLUMEN DE TEMPERATURA

ENERO 1997

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	13.9	14.1	13.6
2	24	14.3	14.5	14.0
3	24	14.5	14.8	14.3
4	24	15.5	15.7	15.3
5	24	15.7	15.9	15.4
6	24	16.6	16.8	16.4
7	24	15.1	15.3	14.9
8	24	15.4	15.5	15.2
9	24	14.7	14.9	14.5
10	24	14.9	15.1	14.8
11	24	12.9	13.1	12.6
12	24	12.0	12.3	11.8
13	24	12.4	12.7	12.1
14	24	11.3	11.5	11.1
15	24	14.8	15.1	14.6
16	24	13.9	14.1	13.7
17	7	6.2	6.2	6.1
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30	17	15.9	16.1	15.7
31	24	14.7	14.9	14.4

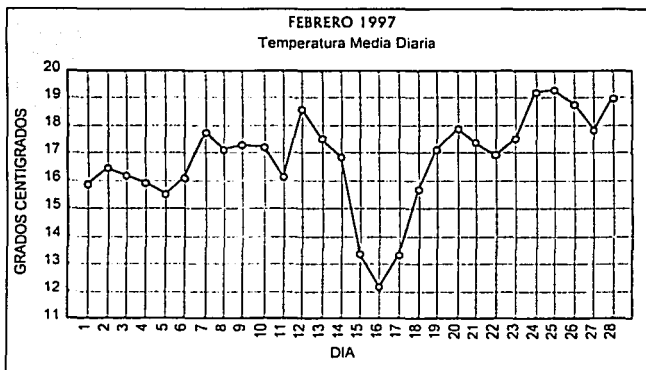
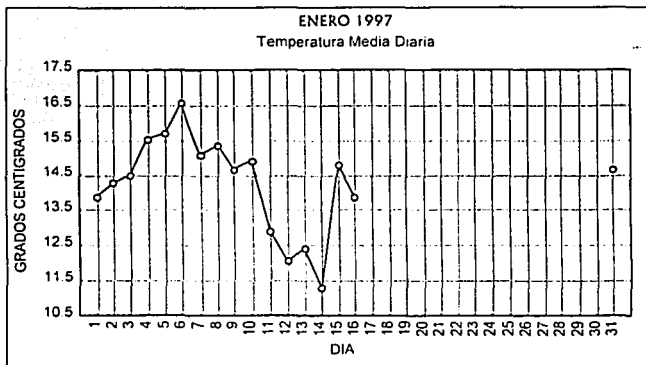
TEMPERATURA MEDIA MENSUAL	14.26*		
TEMPERATURA MÁXIMA EXTREMA EN EL MES	25.7	EL DIA	6 A LAS 14 HRS
TEMPERATURA MÍNIMA EXTREMA EN EL MES	3.1	EL DIA	13 A LAS 7 HRS

* sin considerar los días del 17 al 30 que no contienen las 24 observaciones del día

FEBRERO 1997

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	15.9	16.1	15.6
2	24	16.4	16.6	16.1
3	24	16.2	16.4	15.9
4	24	15.9	16.2	15.7
5	24	15.5	15.7	15.3
6	24	16.1	16.3	15.9
7	24	17.7	17.9	17.5
8	24	17.1	17.3	16.9
9	24	17.3	17.5	17.1
10	24	17.2	17.5	17.0
11	24	16.1	16.3	15.9
12	24	18.6	18.8	18.3
13	24	17.5	17.7	17.3
14	24	16.8	17.1	16.6
15	24	13.4	13.6	13.1
16	24	12.2	12.4	11.9
17	24	13.3	13.5	13.1
18	24	15.7	16.0	15.5
19	24	17.1	17.3	16.9
20	24	17.8	18.1	17.6
21	24	17.4	17.6	17.1
22	24	16.9	17.2	16.7
23	24	17.5	17.7	17.2
24	24	19.2	19.4	19.0
25	24	19.3	19.5	19.0
26	24	18.8	19.0	18.5
27	24	17.8	18.1	17.6
28	24	19.0	19.3	18.8

TEMPERATURA MEDIA MENSUAL	16.77		
TEMPERATURA MÁXIMA EXTREMA EN EL MES	28.2	EL DIA 24 A LAS 16 HRS	
TEMPERATURA MÍNIMA EXTREMA EN EL MES	4.7	EL DIA 17 A LAS 7 HRS	



MARZO 1997

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	21.2	21.4	20.9
2	24	20.1	20.4	19.8
3	24	20.1	20.4	19.8
4	24	20.9	21.1	20.6
5	24	19.5	19.7	19.2
6	24	14.6	14.7	14.4
7	24	16.0	16.2	15.8
8	24	16.4	16.6	16.1
9	24	16.4	16.6	16.1
10	24	16.3	16.6	16.1
11	24	16.5	16.7	16.2
12	24	17.6	17.8	17.4
13	24	18.8	19.1	18.6
14	24	18.8	19.0	18.6
15	24	19.2	19.5	19.0
16	24	18.4	18.6	18.1
17	24	18.0	18.2	17.7
18	24	16.5	16.8	16.2
19	24	13.7	13.9	13.6
20	24	12.3	12.5	12.1
21	24	14.1	14.2	13.9
22	24	13.5	13.7	13.3
23	24	15.5	15.7	15.3
24	24	16.9	17.1	16.7
25	24	18.3	18.5	18.1
26	24	18.7	18.9	18.5
27	24	19.1	19.3	18.9
28	24	18.5	18.8	18.3
29	24	18.9	19.1	18.7
30	24	19.6	19.9	19.4
31	24	18.7	19.0	18.5

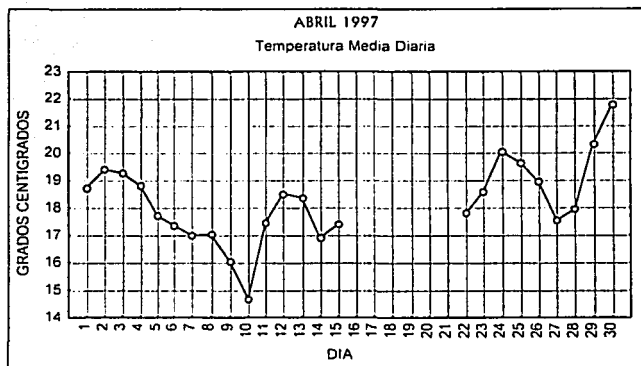
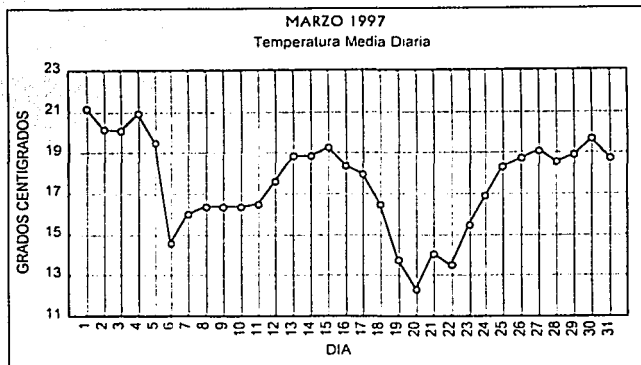
TEMPERATURA MEDIA MENSUAL	17.52	
TEMPERATURA MÁXIMA EN EL MES	30.1	EL DIA 1 A LAS 15 HRS
TEMPERATURA MÍNIMA EN EL MES	7.4	EL DIA 20 A LAS 4 Y 5 HRS

ABRIL, 1997

DIA	No. OBSERVACIONES	TEMPERATURA DIARIA	TEMPERATURA MEDIA	TEMPERATURA MAXIMA	TEMPERATURA MINIMA
1	24	18.7	18.9	18.6	18.6
2	24	19.4	19.7	19.7	19.1
3	24	19.5	19.5	19.6	19.6
4	24	18.8	19.0	19.0	18.6
5	24	17.7	17.9	17.8	17.5
6	24	17.4	17.6	17.6	17.1
7	24	17.0	17.2	17.2	16.7
8	24	17.0	17.5	17.5	16.8
9	24	16.0	16.5	16.5	15.8
10	24	14.7	14.8	14.8	14.6
11	24	17.4	17.7	17.7	17.2
12	24	18.5	18.7	18.7	18.4
13	24	18.4	18.6	18.6	18.2
14	24	16.9	17.1	17.1	16.7
15	24	17.4	17.7	17.7	17.1
16	11	15.4	15.7	15.7	15.2
17					
18					
19					
20					
21	14	18.0	18.5	18.5	17.6
22	24	17.8	18.2	18.2	17.4
23	24	18.6	19.1	19.1	18.1
24	24	20.0	20.6	20.6	19.6
25	24	19.6	20.1	20.1	19.1
26	24	18.9	19.4	19.4	18.5
27	24	17.5	17.9	17.9	17.1
28	24	17.9	18.3	18.3	17.5
29	24	20.3	20.6	20.6	20.0
30	23	21.8	22.1	22.1	21.5

TEMPERATURA MEDIA MENSUAL	18.27
TEMPERATURA MAXIMIA EXTREMIA EN EL MES	29.6
EL DIA 30 A LAS 5 HRS	9.6
TEMPERATURA MINIMIA EXTREMIA EN EL MES	9.6
EL DIA 28 A LAS 5 HRS	

sin considerar los dias del 16 al 21 que no contienen las 24 observaciones del dia, pero si incluye el dia 30



MAYO 1997

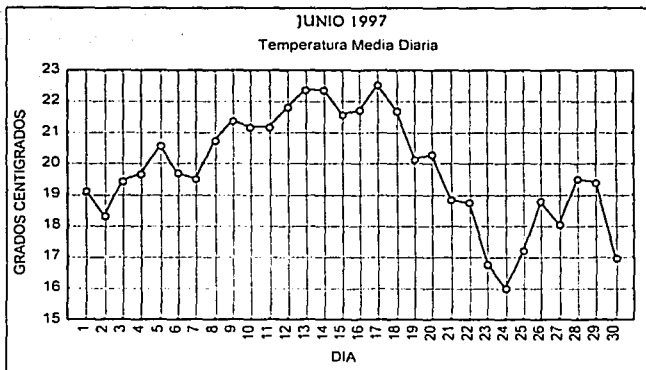
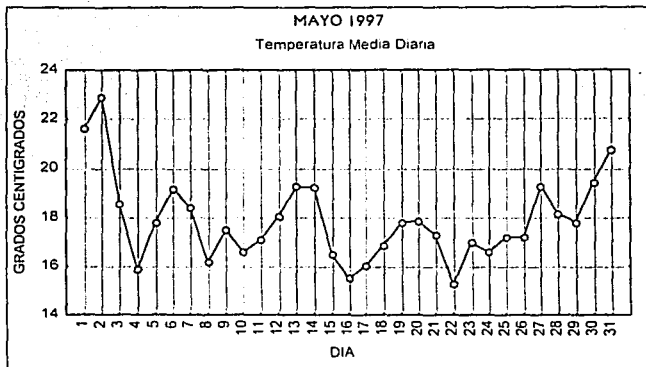
DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	21.6	22.0	21.1
2	24	22.0	23.3	22.5
3	24	18.6	19.1	18.1
4	24	15.9	16.3	15.5
5	24	17.8	18.3	17.3
6	24	19.2	19.7	18.7
7	24	18.4	18.8	17.9
8	24	16.2	16.5	15.8
9	23	17.5	17.9	17.0
10	24	16.6	16.9	16.2
11	24	17.1	17.5	16.7
12	24	18.1	18.6	17.7
13	24	19.3	19.7	18.8
14	24	19.2	19.6	18.9
15	24	16.5	16.9	16.0
16	24	15.5	16.0	15.0
17	24	16.0	16.5	15.6
18	24	16.9	17.4	16.4
19	24	17.8	18.2	17.4
20	24	17.9	18.4	17.4
21	24	17.3	17.7	17.0
22	24	15.3	15.7	14.9
23	24	17.0	17.4	16.5
24	24	16.6	17.0	16.2
25	24	17.2	17.6	16.8
26	24	17.2	17.5	16.8
27	24	19.3	19.7	18.9
28	24	18.2	18.6	17.8
29	24	17.8	18.2	17.5
30	24	19.4	19.8	19.0
31	24	20.8	21.2	20.4

TEMPERATURA MEDIA MENSUAL	17.89	
TEMPERATURA MAXIMA EXTREMA EN EL MES	30.7	EL DIA 2 A LAS 14 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	10.4	EL DIA 9 A LAS 4 Y 5 HRS

JUNIO 1997

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	19.1	19.4	18.8
2	24	18.3	18.8	17.8
3	24	19.4	19.8	19.1
4	24	19.7	20.1	19.3
5	24	20.6	20.9	20.2
6	24	19.7	20.2	19.2
7	24	19.5	19.9	19.1
8	24	20.7	21.1	20.4
9	24	21.4	21.8	21.0
10	24	21.2	21.6	20.8
11	24	21.2	21.7	20.8
12	24	21.8	22.2	21.4
13	23	22.4	22.9	21.9
14	24	22.4	22.8	22.0
15	24	21.6	22.0	21.2
16	24	21.7	22.1	21.4
17	24	22.5	23.0	22.1
18	24	21.7	22.2	21.3
19	24	20.1	20.6	19.8
20	23	20.3	20.7	19.9
21	24	18.8	19.2	18.4
22	24	18.7	19.1	18.3
23	24	16.8	17.1	16.4
24	24	16.0	16.3	15.7
25	24	17.2	17.5	16.9
26	24	18.8	19.3	18.3
27	24	18.0	18.5	17.7
28	24	19.5	19.9	19.0
29	24	19.4	19.8	19.0
30	24	17.0	17.3	16.6

TEMPERATURA MEDIA MENSUAL		19.84
TEMPERATURA MÁXIMA EXTREMA EN EL MES	30.7	EL DÍA 15 A LAS 16 HRS
TEMPERATURA MÍNIMA EXTREMA EN EL MES	12.1	EL DÍA 4 A LAS 6 HRS



JULIO 1997

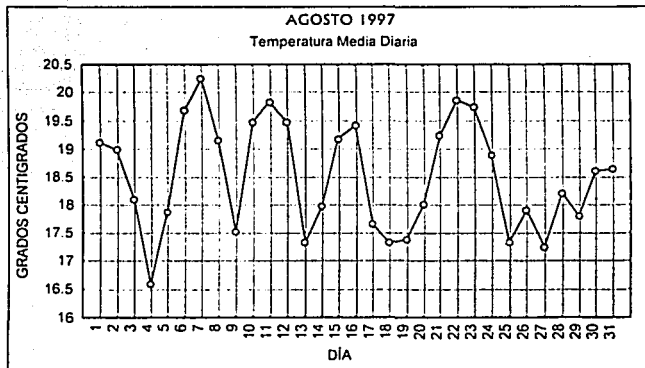
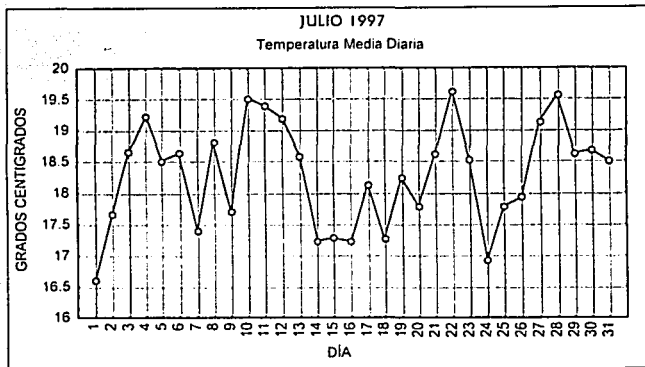
DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	16.6	17.0	16.3
2	24	17.7	18.0	17.3
3	24	18.7	19.2	18.3
4	24	19.2	19.7	18.7
5	24	18.5	18.9	18.2
6	24	18.6	19.1	18.1
7	24	17.4	17.9	16.8
8	24	18.8	19.2	18.5
9	24	17.7	18.1	17.4
10	24	19.5	19.9	19.2
11	24	19.4	19.9	18.8
12	24	19.2	19.6	18.7
13	24	18.6	19.0	18.3
14	24	17.2	17.7	16.9
15	24	17.3	17.7	17.0
16	24	17.2	17.5	16.9
17	24	18.1	18.5	17.8
18	24	17.3	17.5	17.0
19	24	18.2	18.6	17.9
20	24	17.8	18.3	17.3
21	24	18.6	19.0	18.2
22	24	19.6	20.0	19.2
23	24	18.5	18.9	18.2
24	24	16.9	17.5	16.5
25	24	17.8	18.3	17.4
26	24	17.9	18.3	17.4
27	24	19.1	19.7	18.7
28	24	19.6	20.0	19.1
29	24	18.6	19.1	18.0
30	24	18.7	19.2	18.1
31	24	18.5	18.9	18.1

TEMPERATURA MEDIA MENSUAL	18.29	
TEMPERATURA MAXMA EXTREMA EN EL MES	28.4	EL DIA 21 A LAS 15 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	12.3	EL DIA 25 A LAS 6 Y 7 HRS

AGOSTO 1997

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	19.1	19.6	18.6
2	24	19.0	19.4	18.6
3	24	18.1	18.6	17.6
4	24	16.6	16.9	16.2
5	24	17.9	18.4	17.5
6	24	19.7	20.1	19.2
7	24	20.2	20.7	19.8
8	24	19.1	19.7	18.6
9	24	17.5	17.8	17.2
10	24	19.5	19.9	18.9
11	24	19.8	20.3	19.4
12	24	19.5	19.9	19.1
13	24	17.3	17.8	16.9
14	24	18.0	18.3	17.7
15	24	19.2	19.7	18.7
16	24	19.4	19.7	19.1
17	24	17.6	18.0	17.3
18	24	17.3	17.7	17.0
19	24	17.4	17.9	16.9
20	24	18.0	18.4	17.6
21	24	19.2	19.7	18.8
22	24	19.9	20.2	19.5
23	24	19.7	20.3	19.3
24	24	18.9	19.4	18.5
25	24	17.3	17.9	16.8
26	24	17.9	18.4	17.5
27	24	17.2	17.6	16.9
28	24	18.2	18.7	17.8
29	24	17.8	18.2	17.4
30	24	18.6	19.1	18.3
31	24	18.6	19.1	18.1

TEMPERATURA MEDIA MIENSUAL	18.5	
TEMPERATURA MÁXIMA EXTREMA EN EL MES	28.7	DIAS 11 Y 23 A LAS 16 HRS
TEMPERATURA MÍNIMA EXTREMA EN EL MES	11.6	EL DIA 24 A LAS 7 HRS



SEPTIEMBRE 1997

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	18.7	19.3	18.1
2	24	18.2	18.8	17.8
3	24	17.2	17.6	16.9
4	24	16.6	16.8	16.3
5	24	16.6	16.9	16.3
6	24	17.2	17.5	16.9
7	24	17.0	17.4	16.6
8	24	17.5	18.0	17.1
9	24	18.8	19.3	18.3
10	13	17.1	17.2	16.9
11	24	19.1	19.7	18.6
12	24	19.1	19.8	18.6
13	24	18.8	19.3	18.4
14	24	17.0	17.4	16.6
15	24	17.6	18.0	17.2
16	24	19.2	19.6	18.6
17	24	18.9	19.4	18.5
18	24	18.6	19.1	17.9
19	24	17.8	18.4	17.4
20	24	16.9	17.2	16.6
21	24	17.8	18.2	17.4
22	24	18.1	18.5	17.8
23	24	19.2	19.7	18.8
24	24	19.0	19.5	18.6
25	24	16.6	16.9	16.4
26	24	17.8	18.3	17.3
27	24	18.4	18.8	18.1
28	24	20.9	21.3	20.5
29	24	20.0	20.3	19.6
30	24	19.6	20.0	19.2

TEMPERATURA MEDIA MENSUAL		18.2*
TEMPERATURA MÁXIMA EXTREMA EN EL MES	29.1	EL DÍA 28 A LAS 14 HRS
TEMPERATURA MÍNIMA EXTREMA EN EL MES	12.0	EL DÍA 15 A LAS 6 HRS

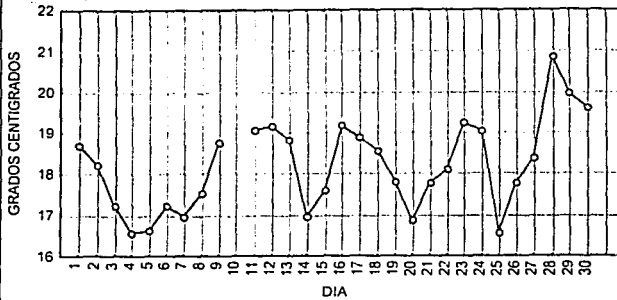
* sin considerar el día 10 que no contiene las 24 observaciones del día

OCTUBRE 1997

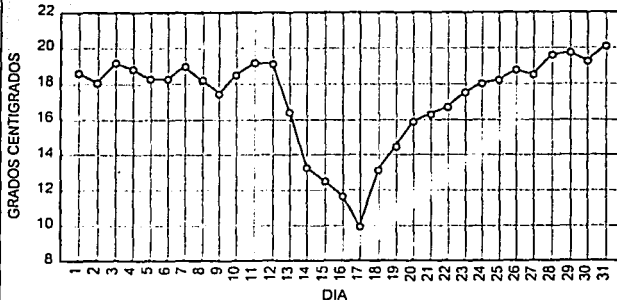
DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	18.6	18.9	18.3
2	24	18.0	18.4	17.7
3	24	19.1	19.5	18.7
4	24	18.8	19.2	18.4
5	24	18.2	18.7	17.9
6	24	18.2	18.6	17.9
7	24	18.9	19.4	18.5
8	24	18.2	18.6	17.9
9	24	17.5	17.7	17.2
10	24	18.5	18.7	18.2
11	24	19.1	19.5	18.8
12	24	19.1	19.5	18.8
13	24	16.4	16.6	16.1
14	24	13.5	13.6	13.0
15	23	12.5	12.7	12.2
16	24	11.7	12.0	11.4
17	24	9.9	10.1	9.7
18	24	13.2	13.6	12.6
19	24	14.5	14.9	14.0
20	24	15.8	16.3	15.3
21	24	16.3	16.7	15.8
22	24	16.6	17.0	16.2
23	24	17.5	17.9	17.1
24	24	18.0	18.4	17.6
25	24	18.2	18.7	17.7
26	24	18.7	19.1	18.3
27	24	18.5	18.8	18.1
28	24	19.6	20.0	19.1
29	24	19.7	20.2	19.3
30	24	19.3	19.7	18.9
31	24	20.1	20.5	19.7

TEMPERATURA MEDIA MENSUAL	17.16	
TEMPERATURA MAXIMA EXTREMA EN EL MES	29.2	EL DIA 28 A LAS 15 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	3.2	EL DIA 18 A LAS 7 HRS

SEPTIEMBRE 1997
Temperatura Media Diaria



OCTUBRE 1997
Temperatura Media Diaria



TEMPERATURA MEDIA MENSUAL	27.7	EL DIA 14 A LAS 7 HRS
TEMPERATURA MAXIMA EXTREMA EN EL MES	27.7	EL DIA 15 A LAS 15 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	7.7	EL DIA 14 A LAS 7 HRS

DIA	No. OBSERVACIONES	TEMPERATURA DIARIA	TEMPERATURA MEDIA MAXIMA	TEMPERATURA MEDIA MINIMA
1	24	20.1	20.5	19.7
2	24	17.0	17.2	16.8
3	24	16.8	17.0	16.5
4	24	17.3	17.6	17.0
5	24	17.8	18.1	17.4
6	24	16.2	16.6	15.7
7	24	16.1	16.5	15.7
8	24	15.8	16.1	15.5
9	24	16.7	17.0	16.5
10	24	17.2	17.6	16.9
11	24	16.3	16.8	16.0
12	24	16.8	17.3	16.4
13	24	17.1	17.6	16.6
14	24	17.3	17.7	16.8
15	24	17.4	17.9	16.9
16	24	15.1	15.5	14.7
17	24	15.3	15.7	14.9
18	24	16.7	17.1	16.4
19	24	16.5	16.9	16.2
20	24	16.5	16.9	16.2
21	24	17.5	17.9	17.3
22	24	17.0	17.3	16.7
23	24	15.3	15.7	15.0
24	24	15.1	15.5	14.8
25	24	15.8	16.2	15.4
26	24	15.2	15.6	14.9
27	24	16.3	16.7	16.0
28	24	16.4	16.8	16.0
29	24	16.6	17.0	16.1
30	24	15.7	16.1	15.2

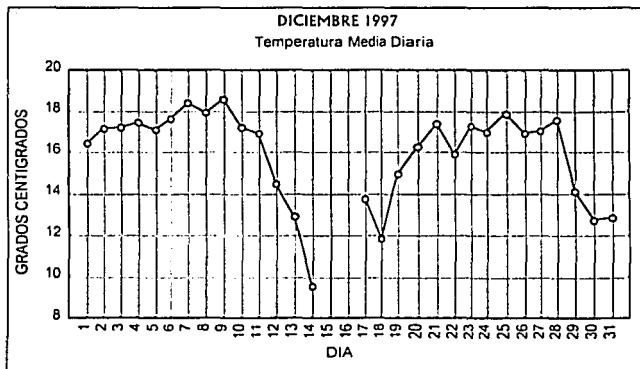
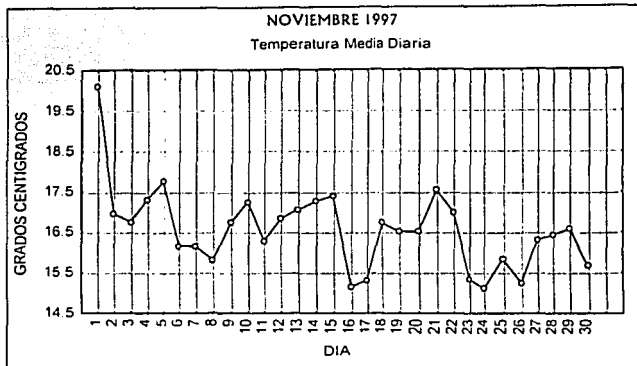
NOVIEMBRE 1997

DICIEMBRE 1997

DIA	Nº. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	16.4	16.9	16.0
2	24	17.2	17.5	16.8
3	24	17.2	17.6	16.8
4	24	17.4	17.8	17.0
5	24	17.1	17.5	16.6
6	24	17.6	18.0	17.3
7	24	18.4	18.8	18.0
8	24	17.9	18.3	17.5
9	24	18.6	19.1	18.1
10	24	17.2	17.7	16.7
11	24	16.9	17.3	16.5
12	24	14.5	14.8	14.2
13	24	12.9	13.2	12.5
14	24	9.6	10.1	9.1
15	15	7.5	8.0	6.9
16	14	16.8	17.2	16.4
17	24	13.7	14.2	13.4
18	24	11.8	12.0	11.6
19	24	14.9	15.2	14.6
20	24	16.3	16.6	15.9
21	24	17.4	17.8	17.1
22	24	15.9	16.2	15.6
23	24	17.2	17.5	16.9
24	24	17.0	17.3	16.7
25	24	17.9	18.3	17.5
26	24	16.9	17.4	16.4
27	24	17.0	17.5	16.6
28	24	17.6	18.1	17.1
29	24	14.1	14.6	13.7
30	24	12.7	13.2	12.3
31	24	12.8	13.4	12.2

TEMPERATURA MEDIA MENSUAL	15.93*	
TEMPERATURA MAXIMA EXTREMA EN EL MES	28.5	EL DIA 9 A LAS 16 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	1.3	EL DIA 15 DE LAS 5-7 HRS

* sin considerar los días 15 y 16 que no contienen las 24 observaciones del día

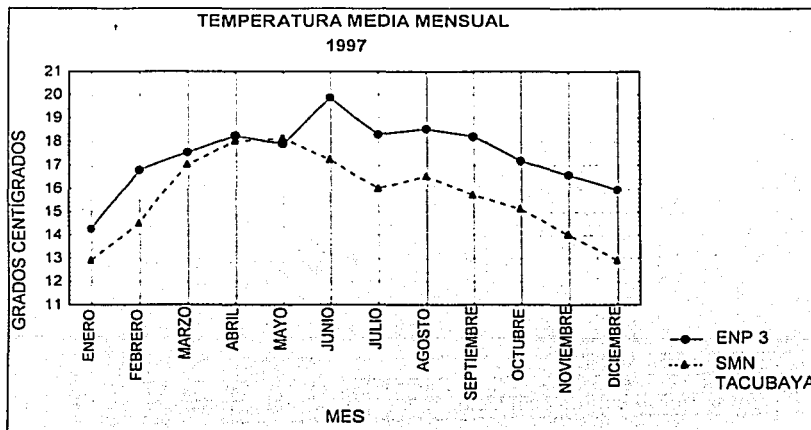


TEMPERATURA MEDIA MENSUAL

1997

	S.M.N. ESTACION TACUBAYA *	E.N.P. PLANTEL N° 3
ENERO	12.9	14.26
FEBRERO	14.5	16.77
MARZO	17.0	17.52
ABRIL	18.0	18.22
MAYO	18.1	17.89
JUNIO	17.2	19.84
JULIO	16.0	18.29
AGOSTO	16.5	18.5
SEPTIEMBRE	15.7	18.2
OCTUBRE	15.1	17.16
NOVIEMBRE	14.0	16.57
DICIEMBRE	12.9	15.93

ANUAL	15.66	17.43
--------------	--------------	--------------



* datos obtenidos en la página del Servicio Meteorológico Nacional

VOLUMEN DE TEMPERATURA

ENERO 1998

DIA	No. OBSERVACIONES	TEMPERATURA EN EL DIA	TEMPERATURA MEDIA	TEMPERATURA MAXIMA	TEMPERATURA MINIMA
1	24	11.4	11.8	11.8	11.0
2	24	12.1	12.4	12.4	11.7
3	24	14.3	14.8	14.8	13.8
4	24	14.6	15.0	15.0	14.2
5	24	14.9	15.4	15.4	14.5
6	24	15.3	15.7	15.7	14.9
7	24	13.6	14.0	14.0	13.3
8	24	13.8	14.3	14.3	13.4
9	24	14.5	14.9	14.9	14.1
10	24	14.9	15.3	15.3	14.4
11	24	15.5	16.0	16.0	15.0
12	24	16.1	16.6	16.6	15.6
13	24	16.0	16.4	16.4	15.6
14	24	15.9	16.4	16.4	15.4
15	24	13.4	13.8	13.8	13.0
16	24	13.0	13.3	13.3	12.6
17	24	14.3	14.5	14.5	14.0
18	24	13.2	13.5	13.5	13.0
19	24	12.7	12.9	12.9	12.4
20	24	14.2	14.7	14.7	13.8
21	24	14.9	15.3	15.3	14.4
22	24	15.4	15.8	15.8	15.0
23	24	14.8	15.4	15.4	14.4
24	24	12.9	13.2	13.2	12.7
25	24	13.3	13.7	13.7	12.8
26	24	13.6	14.1	14.1	13.1
27	24	13.4	13.8	13.8	13.1
28	24	14.5	14.9	14.9	14.1
29	24	15.9	16.2	16.2	15.5
30	24	16.0	16.4	16.4	15.5
31	24	15.5	16.0	16.0	15.0

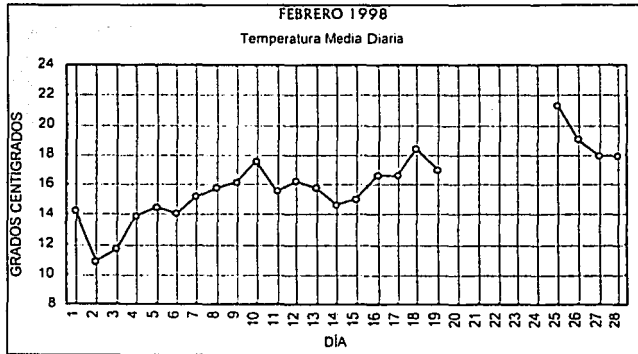
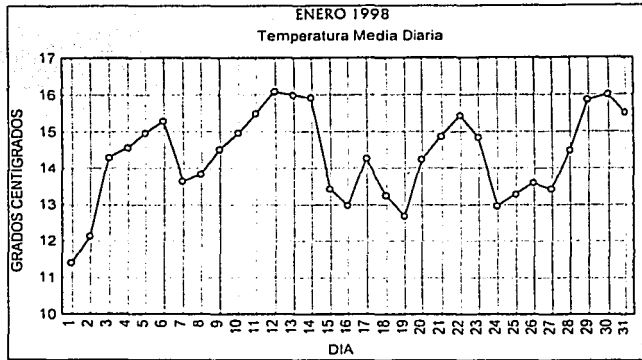
TEMPERATURA MEDIA MENSUAL	14.32
TEMPERATURA MAXIMA EXTREMA EN EL MES	27.3
EL DIA 11 A LAS 15 HRS	
TEMPERATURA MINIMA EXTREMA EN EL MES	4.6
EL DIA 25 A LAS 7-8 HRS	

FEBRERO 1998

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	14.2	14.6	13.8
2	24	10.9	11.2	10.6
3	24	11.6	12.1	11.1
4	24	13.8	14.4	13.3
5	24	14.4	14.8	14.1
6	24	14.0	14.5	13.6
7	24	15.1	15.6	14.7
8	24	15.7	16.2	15.3
9	24	16.1	16.6	15.7
10	24	17.5	18.0	17.1
11	24	15.5	15.9	15.1
12	24	16.2	16.6	15.8
13	24	15.7	16.1	15.3
14	24	14.6	15.0	14.3
15	24	15.0	15.5	14.4
16	24	16.6	17.1	16.1
17	24	16.5	17.4	15.6
18	24	18.4	19.2	17.5
19	24	17.0	18.0	16.0
20	7	9.6	10.2	9.0
21				
22				
23				
24	2	14.4	15.9	13.0
25	17	21.3	21.9	20.8
26	24	19.0	19.6	18.4
27	24	17.9	18.8	17.0
28	24	17.9	18.8	17.0

TEMPERATURA MEDIA MENSUAL	15.62°		
TEMPERATURA MÁXIMA EXTREMA EN EL MES	28.4	EL DIA 17 A LAS 15-16 HRS	
TEMPERATURA MÍNIMA EXTREMA EN EL MES	2.1	EL DIA 3 A LAS 7 HRS	

* sin considerar los días del 20 al 25 que no contienen las 24 observaciones del día.



MARZO 1998

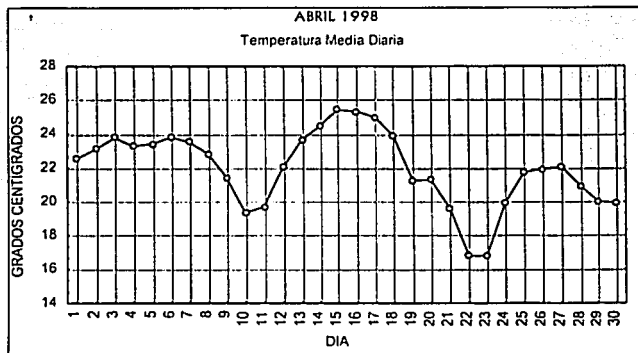
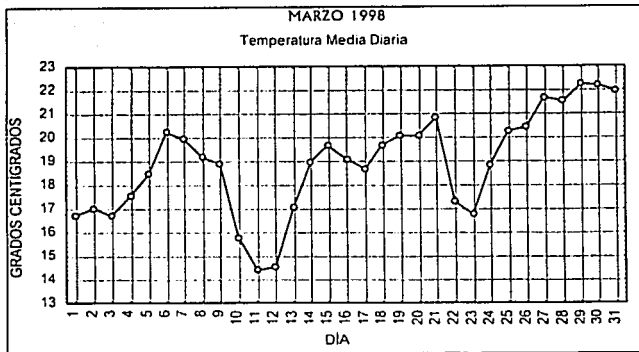
DIA	NO. OBSERVACIONES	TEMPERATURA EN EL DIA	TEMPERATURA DIARIA	TEMPERATURA MEDIA	TEMPERATURA MAXIMA	TEMPERATURA MINIMA
1	24	16.7	17.6	17.6	15.8	15.8
2	24	17.0	18.0	18.0	16.1	16.1
3	24	16.7	17.6	17.6	15.8	15.8
4	24	17.5	18.2	18.2	16.7	16.7
5	24	18.5	19.4	19.4	17.6	17.6
6	24	20.2	21.0	21.0	19.4	19.4
7	24	19.9	20.7	20.7	19.2	19.2
8	24	19.1	20.2	20.2	18.2	18.2
9	24	18.9	19.7	19.7	18.1	18.1
10	24	15.8	16.8	16.8	14.9	14.9
11	24	14.4	15.3	15.3	13.5	13.5
12	24	14.5	15.2	15.2	13.7	13.7
13	24	17.1	18.0	18.0	16.2	16.2
14	24	19.0	19.8	19.8	18.2	18.2
15	24	19.6	20.4	20.4	18.9	18.9
16	24	19.1	19.7	19.7	18.4	18.4
17	24	18.6	19.4	19.4	17.8	17.8
18	24	19.7	20.5	20.5	18.8	18.8
19	24	20.1	20.9	20.9	19.2	19.2
20	24	20.1	20.9	20.9	19.2	19.2
21	24	20.8	21.8	21.8	19.9	19.9
22	24	17.3	18.1	18.1	16.6	16.6
23	24	16.7	17.7	17.7	15.9	15.9
24	24	18.8	19.7	19.7	17.9	17.9
25	24	20.2	21.1	21.1	19.5	19.5
26	24	20.4	21.3	21.3	19.6	19.6
27	24	21.7	22.2	22.2	21.2	21.2
28	24	21.5	21.9	21.9	21.0	21.0
29	24	22.2	23.6	23.6	21.8	21.8
30	24	22.2	23.5	23.5	21.8	21.8
31	24	22.0	22.4	22.4	21.5	21.5

TEMPERATURA MEDIA MENSUAL	18.91
TEMPERATURA MAXIMA EXTREMA EN EL MES	31.6
TEMPERATURA MINIMA EXTREMA EN EL MES	4.7
EL DIA 11 A LAS 7 HRS	
EL DIA 31 A LAS 15 HRS	

ABRIL 1998

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	22.5	23.0	22.2
2	24	23.1	23.5	22.7
3	24	23.8	24.2	23.4
4	24	23.3	23.8	22.8
5	24	23.4	23.9	23.0
6	24	23.8	24.2	23.4
7	24	23.5	23.9	23.2
8	24	22.8	23.2	22.4
9	24	21.4	21.9	21.0
10	24	19.3	19.8	18.9
11	24	19.7	20.2	19.3
12	24	22.0	22.5	21.6
13	24	23.7	24.2	23.2
14	24	24.4	24.9	24.0
15	23	25.4	25.8	25.0
16	23	25.3	26.0	24.5
17	24	25.0	25.7	24.2
18	24	23.9	24.6	23.2
19	24	21.3	22.0	20.6
20	21	21.3	21.9	20.7
21	24	19.6	20.4	19.0
22	24	16.8	17.6	16.1
23	24	16.8	17.5	16.3
24	24	19.9	20.7	19.3
25	24	21.7	22.5	21.1
26	24	21.9	22.7	21.2
27	24	22.1	22.9	21.4
28	24	20.9	21.6	20.3
29	24	20.0	20.8	19.3
30	24	19.9	20.6	19.3

TEMPERATURA MEDIA MENSUAL		21.95
TEMPERATURA MÁXIMA EXTREMA EN EL MES	33.3	EL DIA 15 A LAS 17 HRS
TEMPERATURA MÍNIMA EXTREMA EN EL MES	10.9	DIAS 10 y 23 A LAS 7 HRS



MAYO 1998

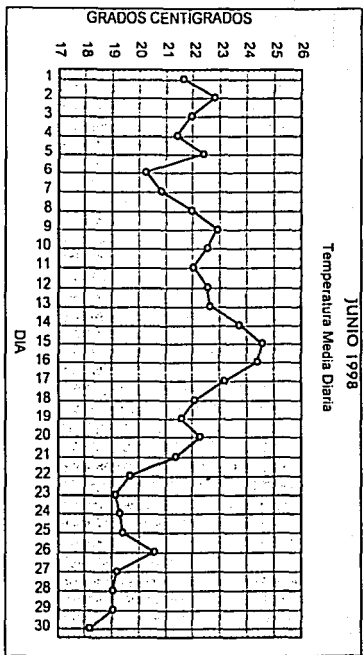
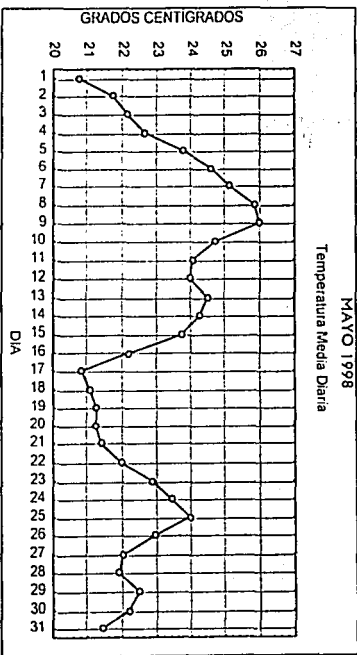
DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MAXIMA	TEMPERATURA MEDIA MINIMA
1	24	20.7	21.5	20.0
2	24	21.7	22.5	20.9
3	24	22.1	23.0	21.3
4	24	22.6	23.4	21.9
5	24	23.8	24.6	23.0
6	24	24.6	25.4	23.8
7	24	25.1	26.0	24.2
8	24	25.8	26.6	25.0
9	24	26.0	26.9	25.2
10	24	24.7	25.7	23.9
11	24	24.0	24.8	23.2
12	24	23.9	24.8	23.3
13	24	24.5	25.3	23.7
14	24	24.2	25.0	23.5
15	24	23.7	24.5	22.9
16	24	22.2	23.2	21.3
17	24	20.8	21.6	20.1
18	24	21.1	22.0	20.3
19	24	21.3	22.2	20.4
20	24	21.2	22.2	20.4
21	24	21.4	22.2	20.6
22	24	22.0	22.7	21.2
23	24	22.9	23.6	22.1
24	24	23.4	24.2	22.7
25	24	23.9	24.7	23.2
26	24	22.9	23.7	22.3
27	24	22.0	22.8	21.3
28	24	21.9	22.7	21.2
29	24	22.5	23.2	21.8
30	24	22.2	22.9	21.5
31	24	21.4	22.3	20.6

TEMPERATURA MEDIA MENSUAL	22.92	
TEMPERATURA MAXIMA EXTREMA EN EL MES	34.8	EL DIA 10 A LAS 15 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	11	EL DIA 20 A LAS 7 HRS

JUNIO 1998

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	21.6	22.4	21.0
2	24	22.8	23.6	22.1
3	24	21.9	22.7	21.2
4	24	21.4	22.2	20.5
5	24	22.4	23.1	21.7
6	24	20.3	21.0	19.4
7	24	20.8	21.6	20.1
8	24	21.9	22.6	21.2
9	24	22.9	23.6	22.1
10	24	22.5	23.2	21.8
11	24	22.0	22.8	21.2
12	24	22.5	23.3	21.7
13	24	22.6	23.4	21.9
14	24	23.7	24.4	23.0
15	24	24.5	25.3	23.7
16	24	24.4	25.0	23.7
17	24	23.1	23.7	22.5
18	24	22.0	22.9	21.1
19	24	21.5	22.3	20.9
20	24	22.3	23.0	21.5
21	24	21.3	22.1	20.6
22	24	19.6	20.4	19.0
23	24	19.1	20.0	18.4
24	24	19.3	20.1	18.5
25	24	19.4	20.1	18.7
26	24	20.5	21.2	19.9
27	24	19.2	19.7	18.5
28	24	19.0	19.7	18.4
29	24	19.0	19.8	18.3
30	24	18.2	18.8	17.6

TEMPERATURA MEDIA MENSUAL	21.39	
TEMPERATURA MAXIMA EXTREMA EN EL MES	33.6	EL DIA 15 A LAS 16 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	13.4	EL DIA 23 A LAS 6-7 HRS



JULIO 1998

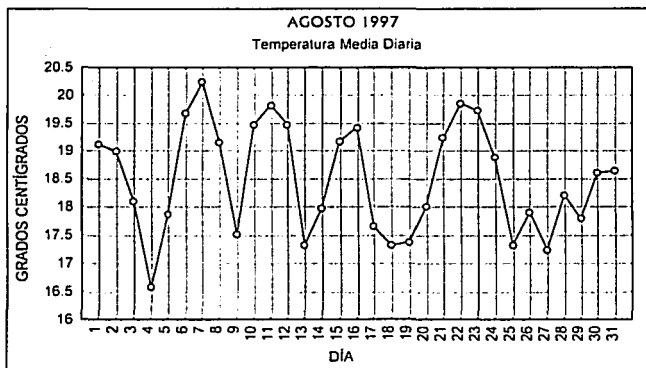
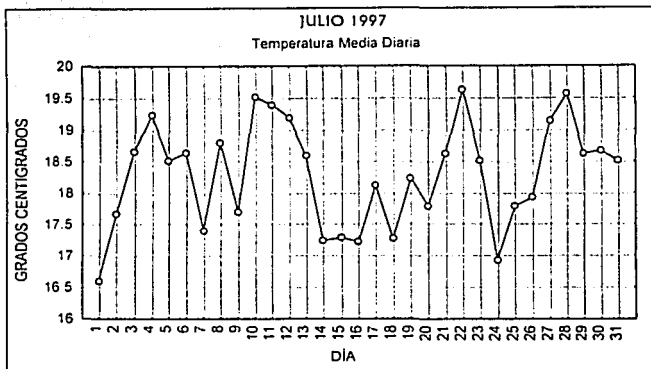
DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	19.0	19.8	18.3
2	24	19.5	20.3	18.8
3	24	17.7	18.3	17.2
4	24	19.9	20.6	19.2
5	24	19.1	19.9	18.4
6	24	19.5	20.1	18.8
7	24	19.8	20.5	19.1
8	24	19.9	20.6	19.2
9	24	19.1	19.7	18.6
10	24	17.5	18.0	17.0
11	24	19.4	20.2	18.7
12	24	20.7	21.6	19.9
13	24	21.4	22.2	20.6
14	24	20.7	21.5	20.1
15	24	19.9	20.6	19.0
16	24	18.9	19.6	18.4
17	24	19.6	20.4	19.0
18	24	19.9	20.8	19.3
19	24	20.1	20.9	19.4
20	24	20.6	21.4	20.0
21	24	20.2	20.9	19.4
22	24	19.2	19.9	18.5
23	24	18.7	19.5	18.0
24	24	19.1	19.9	18.5
25	24	18.8	19.5	18.0
26	24	18.6	19.2	18.0
27	24	18.7	19.5	17.9
28	24	17.9	18.7	17.2
29	24	18.0	18.7	17.2
30	24	19.5	20.3	18.8

TEMPERATURA MEDIA MENSUAL	19.32	
TEMPERATURA MÁXIMA EXTREMA EN EL MES	30.3	DÍAS 12-13 A LAS 17 y 16 HRS
TEMPERATURA MÍNIMA EXTREMA EN EL MES	12.2	EL DIA 25 A LAS 7 HRS

AGOSTO 1998

DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	18.1	19.0	17.4
2	24	18.3	19.0	17.6
3	24	18.8	19.4	18.2
4	24	18.3	19.0	17.9
5	24	18.1	19.0	17.4
6	24	17.9	18.5	17.1
7	24	17.9	18.4	17.5
8	24	19.1	19.8	18.5
9	24	18.2	19.1	17.3
10	24	17.1	17.7	16.5
11	24	17.5	18.3	16.8
12	24	17.8	18.7	17.0
13	24	18.9	19.8	18.2
14	23	18.7	19.6	18.1
15	24	19.0	19.8	18.3
16	24	17.9	18.4	17.2
17	24	18.8	19.6	18.1
18	24	19.7	20.4	19.0
19	24	18.5	19.3	18.0
20	24	16.3	16.9	15.9
21	24	17.6	18.4	16.9
22	24	19.1	19.8	18.6
23	24	19.7	20.4	19.0
24	24	19.9	20.8	19.2
25	24	19.6	20.5	18.9
26	24	19.1	19.9	18.4
27	21	19.3	20.0	18.7
28	24	18.9	19.5	18.3
29	24	18.9	19.5	18.3
30	24	18.3	18.9	17.7
31	23	18.1	18.7	17.5

TEMPERATURA MEDIA MENSUAL		18.5
TEMPERATURA MAXIMA EXTREMA EN EL MES	28.2	EL DIA 13 A LAS 16 HRS
TEMPERATURA MINIMA EXTREMA EN EL MES	11.3	EL DIA 10 A LAS 7 HRS



SEPTIEMBRE 1998

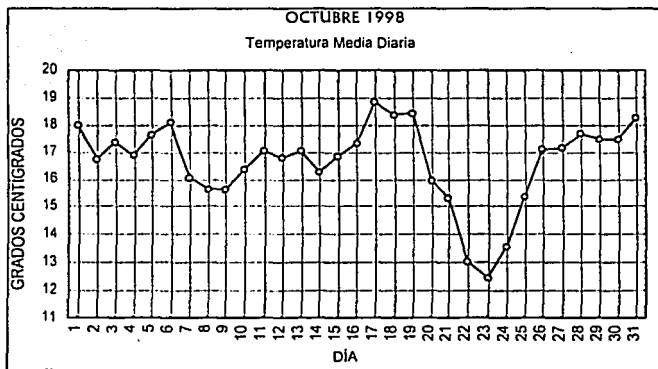
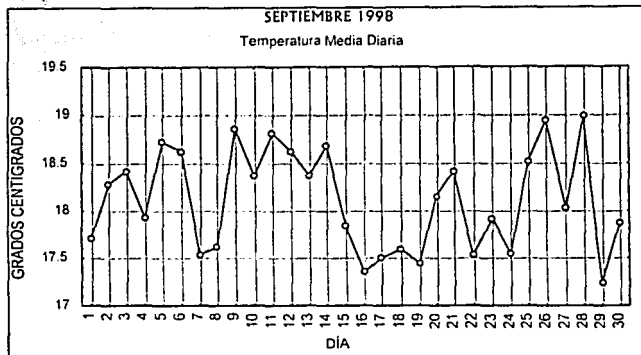
DIA	No. OBSERVACIONES EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MÁXIMA	TEMPERATURA MEDIA MÍNIMA
1	24	17.7	18.3	17.3
2	24	18.3	19.0	17.5
3	24	18.4	19.0	17.9
4	24	17.9	18.4	17.4
5	24	18.7	19.3	18.1
6	24	18.6	19.2	18.1
7	24	17.5	18.2	17.0
8	24	17.6	18.1	17.2
9	24	18.9	19.5	18.4
10	23	18.4	19.0	17.8
11	24	18.8	19.5	18.3
12	24	18.6	19.1	18.1
13	24	18.4	18.9	17.9
14	24	18.7	19.4	18.1
15	24	17.8	18.3	17.4
16	24	17.4	17.9	16.9
17	23	17.5	18.2	16.9
18	24	17.6	18.2	17.0
19	24	17.4	18.0	17.0
20	24	18.1	18.7	17.5
21	24	18.4	18.9	17.9
22	24	17.5	18.2	16.9
23	24	17.9	18.5	17.4
24	24	17.5	18.0	17.1
25	24	18.5	19.2	18.0
26	24	18.9	19.6	18.3
27	24	18.0	18.6	17.4
28	24	19.0	19.7	18.4
29	24	17.2	17.9	16.8
30	24	17.9	18.5	17.4

TEMPERATURA MEDIA MENSUAL	18.10		
TEMPERATURA MÁXIMA EXTREMA EN EL MES	26.7	DÍAS 10 Y 25 A LAS 13 Y 14 HRS	
TEMPERATURA MÍNIMA EXTREMA EN EL MES	13.3	EL DÍA 4 A LAS 5 HRS	

OCTUBRE 1998

DIA	Nº. OBSERVACIONES	EN EL DIA	TEMPERATURA MEDIA	TEMPERATURA MAXIMA	TEMPERATURA MINIMA
1	24	18.0	18.7	18.7	17.6
2	24	16.7	17.4	17.4	16.2
3	24	17.4	17.9	17.9	16.8
4	24	16.9	17.5	17.5	16.3
5	24	17.6	18.4	18.4	17.0
6	24	18.1	18.7	18.7	17.6
7	24	16.0	16.7	16.7	15.5
8	24	15.6	16.0	16.0	15.3
9	24	15.6	16.0	16.0	15.2
10	24	16.3	16.8	16.8	15.8
11	24	17.0	17.5	17.5	16.7
12	24	16.8	17.2	17.2	16.4
13	24	17.0	17.6	17.6	16.6
14	24	16.3	16.8	16.8	15.7
15	24	16.8	17.4	17.4	16.3
16	23	17.3	17.9	17.9	16.7
17	24	18.8	19.6	19.6	18.0
18	23	18.4	19.2	19.2	17.6
19	24	18.4	19.0	19.0	17.7
20	24	15.9	16.9	16.9	15.3
21	24	15.3	15.9	15.9	14.9
22	24	13.0	13.4	13.4	12.7
23	24	12.4	12.8	12.8	12.1
24	24	13.6	14.0	14.0	13.1
25	24	15.3	15.8	15.8	15.0
26	24	17.1	17.9	17.9	16.4
27	24	17.1	17.9	17.9	16.4
28	24	17.6	18.5	18.5	16.9
29	24	17.5	18.3	18.3	16.7
30	24	17.4	18.3	18.3	16.5
31	24	18.2	19.0	19.0	17.6

TEMPERATURA MEDIA MENSUAL	16.63
TEMPERATURA MAXIMA EXTREMA EN EL MES	27.3
TEMPERATURA MINIMA EXTREMA EN EL MES	7.8
EL DIA 30 A LAS 6-7 HRS	

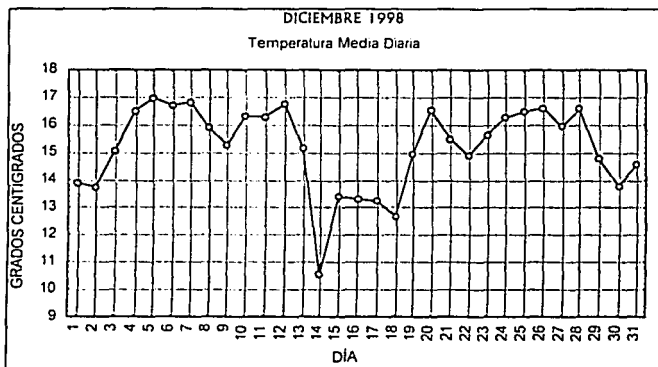
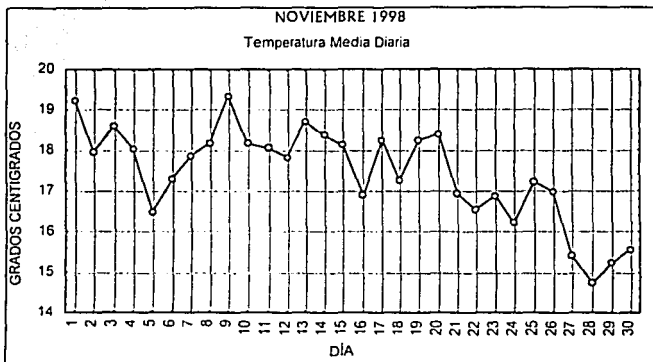


TEMPERATURA MEDIA MENSUAL	17.43
TEMPERATURA MAXIMIA EXTREMA EN EL MES	29.4
TEMPERATURA MINIMIA EXTREMA EN EL MES	6.8
EL DIA 28 A LAS 7-8 HRS	

DIA	NO. OBSERVACIONES	TEMPERATURA EN EL DIA	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MAXIMA	TEMPERATURA MEDIA MINIMA
1	24	19.2	20.1	18.3	18.3
2	24	18.0	18.8	17.2	17.2
3	24	18.6	19.3	17.9	17.9
4	24	18.0	18.7	17.3	17.3
5	24	16.5	17.3	15.8	15.8
6	24	17.3	18.0	16.6	16.6
7	24	17.8	18.6	17.0	17.0
8	24	18.2	18.9	17.4	17.4
9	24	19.3	20.3	18.3	18.3
10	24	18.2	19.1	17.3	17.3
11	24	18.0	18.9	17.2	17.2
12	24	17.8	18.3	17.2	17.2
13	24	18.7	19.6	17.9	17.9
14	24	18.4	19.2	17.6	17.6
15	24	18.1	18.8	17.3	17.3
16	24	16.9	17.5	16.4	16.4
17	24	18.2	19.1	17.5	17.5
18	24	17.2	18.2	16.4	16.4
19	24	18.2	19.2	17.4	17.4
20	24	18.4	19.4	17.7	17.7
21	24	16.9	17.7	16.2	16.2
22	24	16.5	17.3	15.8	15.8
23	24	16.9	17.6	16.2	16.2
24	24	16.2	16.7	15.8	15.8
25	24	17.2	18.0	16.6	16.6
26	24	17.0	17.7	16.3	16.3
27	24	15.4	16.2	14.7	14.7
28	24	14.7	15.6	13.9	13.9
29	24	15.2	16.2	14.4	14.4
30	24	15.6	16.5	14.8	14.8

TEMPERATURA MEDIA MENSUAL	15.18
TEMPERATURA MAXIMA EXTREMA EN EL MES	28.10
TEMPERATURA MINIMA EXTREMA EN EL MES	4.1
EL DIA 18 A LAS 6 HRS	

DIA	No. OBSERVACIONES	TEMPERATURA MEDIA DIARIA	TEMPERATURA MEDIA MAXIMA	TEMPERATURA MEDIA MINIMA
1	24	13.9	14.8	13.1
2	24	13.7	14.6	12.9
3	24	13.0	16.0	14.2
4	24	16.5	17.4	15.7
5	24	17.0	18.0	16.0
6	24	16.7	17.6	15.9
7	24	16.8	18.0	16.0
8	24	15.9	17.0	15.0
9	24	15.3	16.2	14.5
10	24	16.3	17.3	15.5
11	24	16.3	17.1	15.4
12	24	16.7	17.6	15.9
13	24	15.1	16.0	14.4
14	24	10.5	10.9	10.1
15	24	13.4	14.1	12.6
16	24	13.3	14.3	12.4
17	24	13.2	14.2	12.4
18	24	12.7	13.6	11.8
19	24	14.9	15.9	14.1
20	24	16.5	17.6	15.5
21	24	15.3	16.7	14.4
22	24	14.9	15.9	13.9
23	24	15.6	16.6	14.9
24	24	16.2	17.2	15.4
25	24	16.3	17.4	15.6
26	24	16.6	17.7	15.6
27	24	15.9	17.0	15.1
28	24	16.5	17.5	15.7
29	24	14.7	15.7	13.9
30	24	13.8	14.8	12.9
31	24	14.5	15.5	13.6

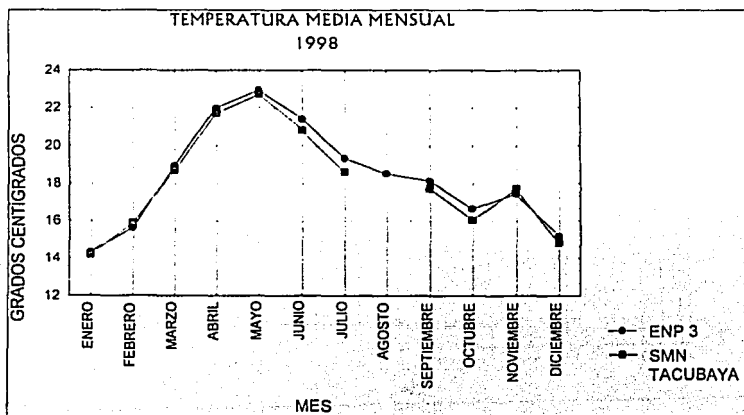


TEMPERATURA MEDIA MENSUAL

1998

	S.M.N. ESTACION TACUBAYA	E.N.P. PLANTEL N° 3
ENERO	14.2	14.32
FEBRERO	15.9	15.62
MARZO	18.7	18.91
ABRIL	21.7	21.95
MAYO	22.7	22.92
JUNIO	20.8	21.39
JULIO	18.6	19.32
AGOSTO		18.50
SEPTIEMBRE	17.7	18.10
OCTUBRE	16.0	16.63
NOVIEMBRE	17.7	17.43
DICIEMBRE	14.8	15.18

ANUAL 18.07 18.36



datos obtenidos en la página del Servicio Meteorológico Nacional

MANUAL

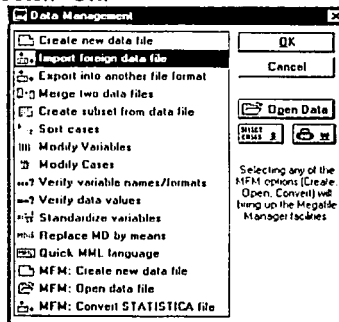
En esta sección se detalla el proceso que realiza cada macro que ha sido mencionado en el documento central. El usuario puede ver aquí lo que su macro hace paso a paso para lograr cada uno de los procesos para los que fue creado. Los macros deben ser ejecutados en el módulo indicado.

También están explicados procesos como la elaboración de gráficas en dos dimensiones.

PROCESO: Importación

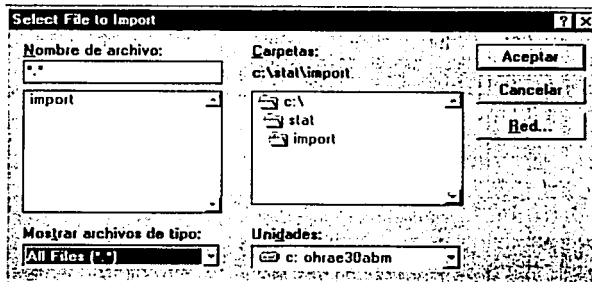
MACRO: Ctrl + I

Paso 1. En el módulo Data Management, pulsa la combinación de teclas Ctrl. + S y aparece el siguiente menú, en él selecciona la segunda opción para la importación de un archivo y oprime el botón OK.



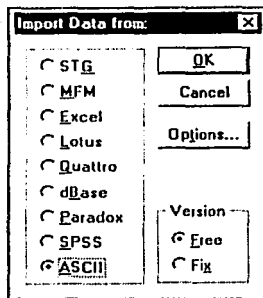
M 1. Panel inicial del módulo Data Management de Statistica. Importación

Paso 2. Elige el archivo a importar c:\stat\import\import



M 2. Diálogo para seleccionar archivos

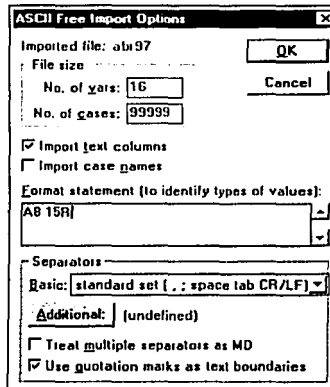
Paso 3. Statistica detecta el tipo de archivo (ASCII) y muestra la siguiente pantalla; presiona el botón OK.



M 3. *Formatos factibles de ser importados por Statistica*

Paso 4. Presenta el siguiente cuadro de diálogo, en donde establecemos que:

- En el tamaño del archivo: son 16 variables y el número de casos lo dejamos intacto
- Por default acepta que se importen las columnas con texto y como en el archivo los casos no tienen asignado un nombre, queda de-seleccionada la siguiente opción
- En el cuadro del centro define que la primera columna de datos, correspondiente a la variable de fecha tiene un formato de 8 posiciones de tipo alfabético y las 15 variables (columnas) restantes son de tipo real
- También define en el último renglón que utilice como separador las comillas. Presiona botón OK.



M 4. *Diálogo para seleccionar opciones para importar archivos ASCII*

Debido a que no especificamos el número exacto de casos (renglones) que contenga el archivo de origen despliega un mensaje que reza "WARNING: TOO FEW ROWS FOUND IN INPUT FILE" y un botón que dice "ACEPTAR", al que da un clic.

La barra de status indica el avance en el proceso de importación, también es señalado con un mensaje que ha sido transformado a un archivo intermedio y posteriormente a uno propio de Statistica, este proceso emplea algunos segundos, preguntando finalmente - mediante un cuadro similar al de la Figura M 2 -, en dónde y con qué nombre salvar el archivo resultante. En este momento termina el macro y el usuario puede elegir el nombre adecuado para su archivo y en dónde desea salvarlo.

PROCESO: Adecuar el archivo importado MACRO: Ctrl + A

Resumen de acciones a realizar en el archivo importado por Statistica. Utilizar el módulo *Data Management*.


ACCIÓN	INCISO
Borra el primer caso que contiene la leyenda "fecha"	1
Elimina la variable "período" y modifica el formato que presenta la "fecha"	2 y 3
Asigna valores texto a la variable "dirección del viento"	4
Define nombre y formato para cada variable	5
Elige solo las observaciones horarias	6
Asigna nombres a los casos para cada día del mes	7

1) Borra el primer caso. Es decir todo el primer renglón (caso).

Case	Fecha	VAR3	VAR4
1	01.01.97	16.400	11.600
2	01.01.97	16.400	11.200
3	01.01.97	16.300	11.100
4	01.01.97	16.300	10.900
5	01.01.97	16.300	10.600
6	01.01.97	16.300	10.200
7	01.01.97	16.300	10.200

M 5. Proceso para borrar un caso

el cuadro de la derecha permite confirmar qué número de caso se desea borrar.

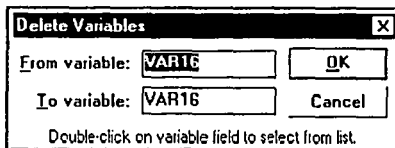
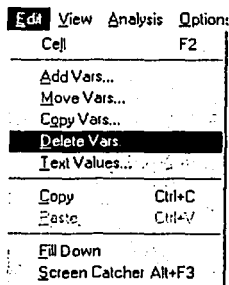
- 2) Escoge el segundo botón de la barra de herramientas (ALL SPECS)  y despliega el siguiente cuadro, en donde es definida cada variable (nombre, código de datos perdidos, formato, nombre largo).

Var	Nombre	HD Code	Format	Long Name: (label, formula or link)
1	VAR1	-9999	8.3	
2	VAR2	-9999	8.3	=v2/100
3	VAR3	-9999	8.3	
4	VAR4	-9999	8.3	
5	VAR5	-9999	8.3	
6	VAR6	-9999	8.3	
7	VAR7	-9999	8.3	
8	VAR8	-9999	8.3	
9	VAR9	-9999	8.3	
10	VAR10	-9999	8.3	
11	VAR11	-9999	8.3	
12	VAR12	-9999	8.3	
13	VAR13	-9999	8.3	
14	VAR14	-9999	8.3	
15	VAR15	-9999	8.3	
16	VAR16	-9999	8.3	

M 6. Cuadro que presenta la descripción de todas las variables


En la variable VAR2 se define como fórmula " $=V2/100$ " para convertir las horas a un formato correcto.

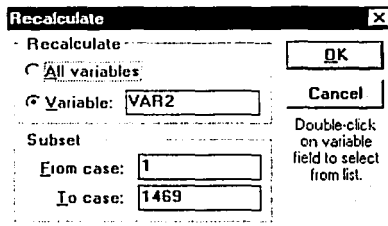
Y ahí mismo borra la variable VAR16 que no tiene uso puesto que solo tiene el período que transcurre de una observación a otra, esto se realiza de la siguiente manera:



M 7. Proceso para borrar una variable desde el cuadro ALL SPECS

como es correcta la selección de la variable a borrar, acepta presionando el botón OK. Luego, cierra la ventana con Ctrl+F4.

- 3) Recalcula la variable VAR2, colocándose en cualquier celda de esa columna y presionando el icono , despliega el siguiente cuadro



Recalculate [X]

Recalculate:

All variables

Variable:

Subset:

From case:

To case:


OK

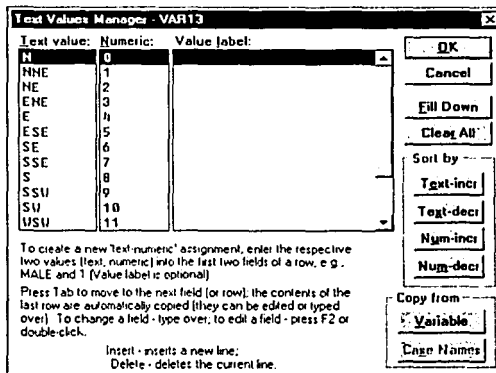
Cancel

Double-click on variable field to select from list.

M 8. Cuadro para recalcular los valores de las variables

en donde acepta que realice el proceso.

- 4) Señala la columna de la VAR13, hace un clic en el botón  y asigna los valores texto correspondientes a la dirección del viento (ver Tabla III) de la siguiente forma¹:



Text Values Manager - VAR13 [X]

Text value: Numeric: Value label:

N	0	
NNE	1	
NE	2	
ENE	3	
E	4	
ESE	5	
SE	6	
SSE	7	
S	8	
SSU	9	
SV	10	
VSU	11	

OK

Cancel

Fill Down

Clear All

Sort by:

Text-incr

Text-decr

Num-incr

Num-decr

Copy from:

Variable

Case Names

To create a new 'Text/numeric' assignment, enter the respective two values (text, numeric) into the first two fields of a row, e.g. MALE and 1 (Value label is optional)

Press Tab to move to the next field (or row); the contents of the last row are automatically copied (they can be edited or typed over). To change a field - type over; to edit a field - press F2 or double-click.

Insert - inserts a new line;
Delete - deletes the current line.

M 9. Asignación de valores texto-número para la dirección del viento

- 5) Define el formato de cada variable, copiando el cuadro de ALL SPECS (ver Inciso 2), de un archivo que ya lo tiene establecido (C:\stat\import\modelo.sta nos

¹ Para pasar de un renglón a otro utilizar el tabulador.

TESIS CCN
FALLA DE ORIGEN

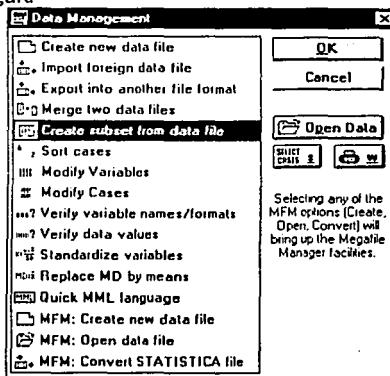
servirá de archivo modelo). Como Statistica no permite abrir dos archivos al mismo tiempo en el mismo módulo, es necesario abrir dicho archivo en otro módulo, por ejemplo en el "Basic Statistics and Tables" y copia el cuadro siguiente

Var	Label	MD Code	Format	Long Name (label, fc)
1	FECHA	-9999	DATE6	
2	HORA	-9999	4 0	HORARIO DE MUESTRA
3	T_INT	-9999	6 1	TEMPERATURA EN EL INTERIOR DE LA I
4	T_EXT	-9999	6 1	TEMPERATURA EN EL EXTERIOR DE LA I
5	T_MAX	-9999	6 1	TEMPERATURA MAXIMA EN EL PERIODO
6	T_MIN	-9999	6 1	TEMPERATURA MINIMA EN EL PERIODO
7	PRES_ATM	-9999	8 1	PRESION ATMOSFERICA
8	HUM_INT	-9999	7 0	HUMEDAD EN EL INTERIOR DE LA ESTAC
9	HUM_EXT	-9999	7 0	HUMEDAD EN EL EXTERIOR DE LA ESTAC
10	P_ROCIO	-9999	7 1	PUNTO DE ROCIO
11	VEL_VTO	-9999	7 1	VELOCIDAD DEL VIENTO
12	RACHA	-9999	6 1	VELOCIDAD MAXIMA DEL VIENTO EN EL
13	DIR_VTO	-9999	7 0	DIRECCION DEL VIENTO
14	ENF_VTO	-9999	7 1	FACTOR DE ENFRIAMIENTO POR EL VIE
15	LLUVIA	-9999	6 0	PRECIPITACION PLUVIAL

M 10. Cuadro para definición de todas las variables (ALL SPECS)

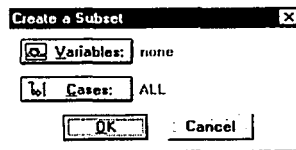
Para después pegarlo en el cuadro (ALL SPECS) del archivo que estamos manejando.

- 6) Como los archivos correspondientes a 1997 y 1998 registraban las observaciones cada 15 o 30 minutos, fue necesario crear un subconjunto del archivo original, que solo tuviera las muestras horarias y eliminara la duplicidad de la hora 24 con la 0 del día siguiente. Con Ctrl + S aparece las siguientes opciones y se escoge la señalada en esta figura



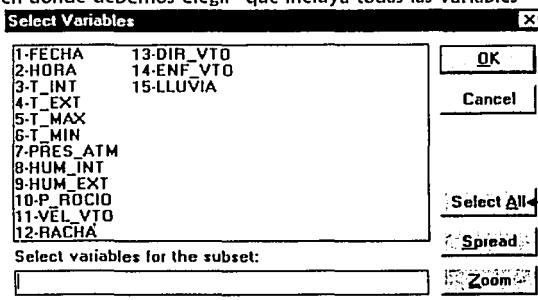
M 11. Panel inicial de módulo Data Management. "Crear subconjunto"

y aparece el cuadro para establecer las condiciones del subconjunto deseado



M 12. Diálogo para definir las características del subconjunto de datos

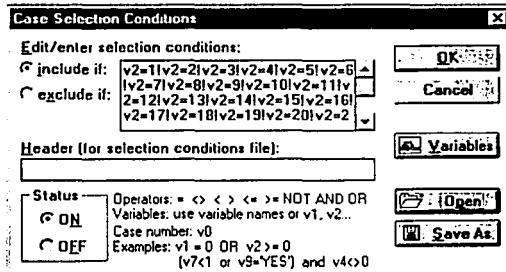
en donde debemos elegir que incluya todas las variables



y presiona OK.

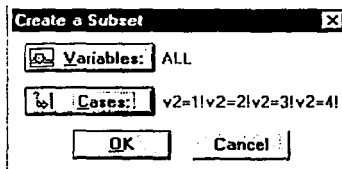
M 13. Diálogo para la selección de variables

Para elegir qué casos requerimos, selecciona los valores en que la variable HORA o V2 toma los valores enteros del período [1...24]. Y define lo que aparece en el cuadro que sigue (el signo "!" que aparece es el operador lógico OR).



M 14. Diálogo para la selección de condiciones

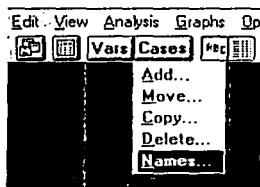
Esta definición puede ser guardada y usada posteriormente si es necesario, mediante la opción "SAVE AS" en el mismo cuadro anterior. Presionar el botón OK. Y finalmente aparece el cuadro, con las opciones ya elegidas



M 15. Diálogo con la definición del subconjunto de datos

como son correctas acepta con OK.

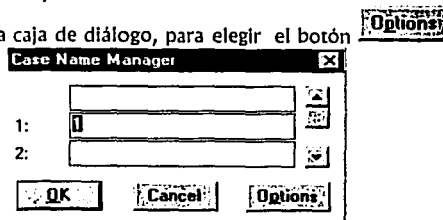
- 7) Pasa al módulo alterno, en donde se encuentra el archivo guía y elige el botón **Cases** en donde aparece el menú de opciones



y elige la opción adecuada (NAMES)

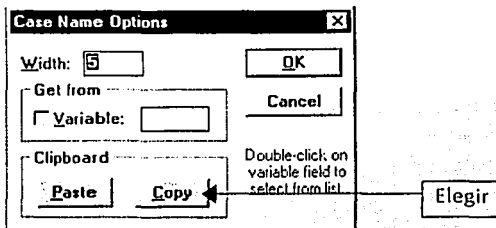
M 16. Menú para el manejo de casos

aparece a continuación la caja de diálogo, para elegir el botón



M 17. Diálogo para asignar nombre a los casos

que a su vez nos transporta al siguiente cuadro, para elegir copiar al portapapeles (clipboard) los nombres asignados

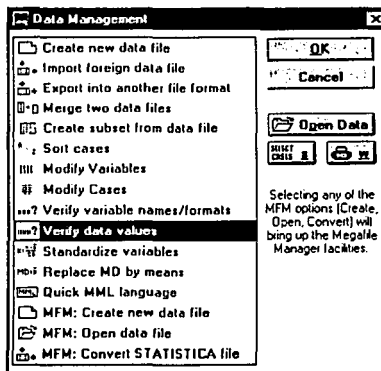


M 18. Diálogo para copiar al portapapeles nombres de los casos


Regresa al archivo que estamos modificando y realiza el proceso de igual manera excepto que al final hay que elegir el botón **Paste**.

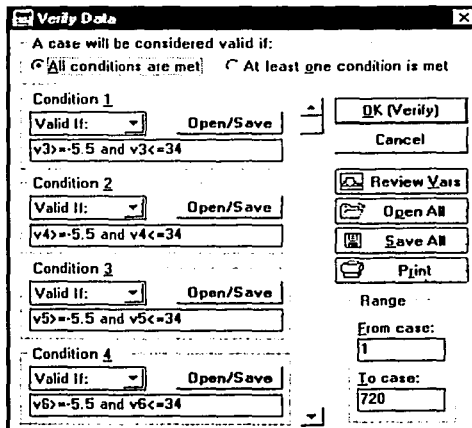
PROCESO: Validación de datos MACRO: Ctrl + B

- 1) Utilizando el módulo Data Management, con la combinación de teclas Ctrl. + S, aparece el siguiente cuadro de diálogo y utilizamos la opción correspondiente para verificar datos



M 19. Cuadro para elegir la verificación de datos

y aparece el siguiente cuadro de diálogo, en el que mediante el botón  **Open All** abrimos el archivo LIMITES.TXT que contiene los valores de tolerancia a verificar y aceptamos con el botón OK(Verify)..



Verify Data

A case will be considered valid if:

All conditions are met At least one condition is met

Condition 1

Valid If: Open/Save

$v3 > -5.5$ and $v3 <= 34$

Condition 2

Valid If: Open/Save

$v4 > -5.5$ and $v4 <= 34$

Condition 3

Valid If: Open/Save

$v5 > -5.5$ and $v5 <= 34$

Condition 4

Valid If: Open/Save

$v6 > -5.5$ and $v6 <= 34$

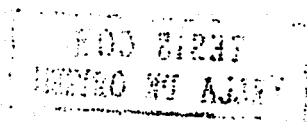
Range

From case:

To case:

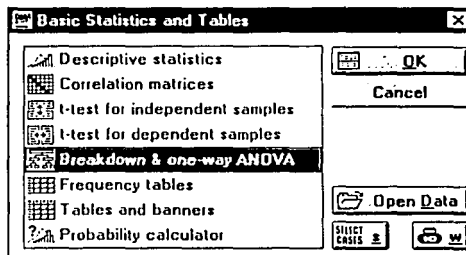
M 20. Cuadro para especificar condiciones de verificación (validación)

Si existe un valor que exceda los límites, el programa indica el número de caso y condición que es excedida, permitiendo editar el dato, ignorar la advertencia o terminar el proceso de verificación.



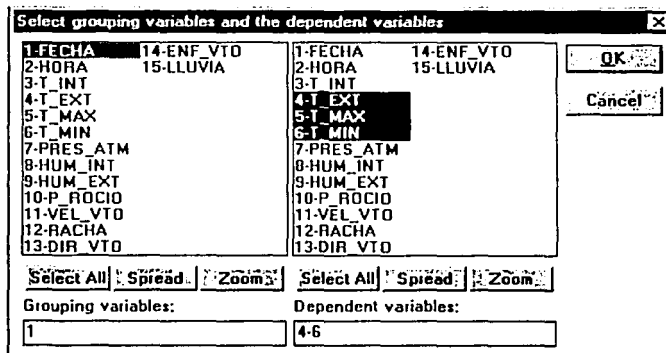
PROCESO: Obtención de Temperatura Media Diaria
MACRO: Ctrl + E

- 1) Utilizando el módulo *Basic Statistics and Tables*, con la combinación de teclas Ctrl. + S, aparece el siguiente cuadro de diálogo y utilizamos la quinta opción.



M 21. Panel inicial del módulo *Basic Statistics and Tables*

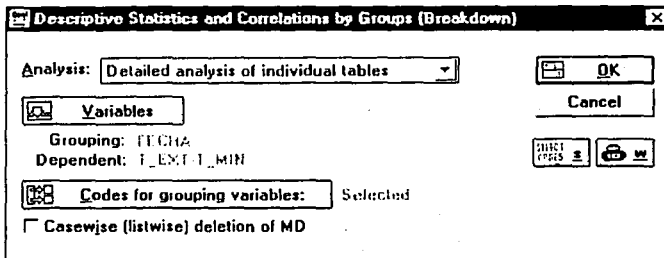
- 2) Presiona el botón OK y aparece la pantalla M 23, es necesario especificar las variables de interés por lo que al presionar el botón "Variables", selecciona las que requerimos, como se muestra



M 22. Cuadro para la selección de variables

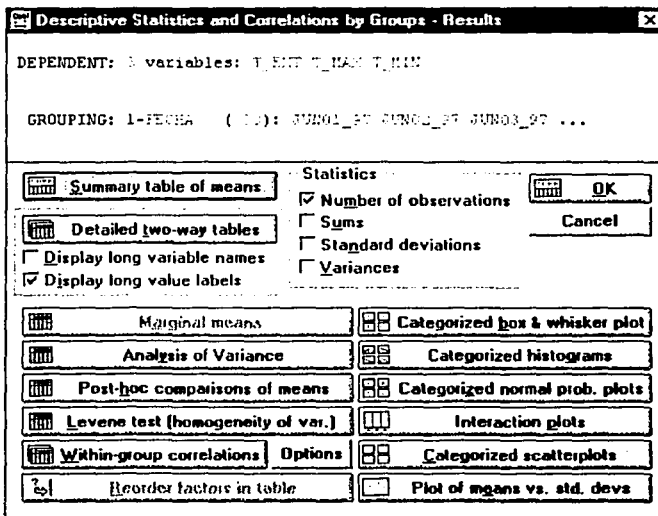
**TESIS CON
FALLA DE ORIGEN**

3) Oprime OK y regresa al diálogo que se presenta a continuación, en donde reporta las variables ya seleccionadas



M 23. Indica la elección de variables para la tabla

4) Elige OK y aparece el cuadro de diálogo en donde con un clic ya sea en OK o en "Summary table of mean" se realizara el calculo deseado.



M 24. Diálogo para elegir la tabla de medios

5) Crea la tabla que reporta las medias de los valores correspondientes a las temperaturas medias diarias y las medias de la temperatura máxima y mínima.

FECHA	T_EXT Mean	T_EXT N	T_MAX Mean	T_MAX N	T_MIN Mean	T_MIN N
JUN01_97	19.108	24	19.383	24	18.821	24
JUN02_97	18.304	24	18.775	24	17.808	24
JUN03_97	19.438	24	19.804	24	19.096	24
JUN04_97	19.675	24	20.083	24	19.258	24
JUN05_97	20.567	24	20.896	24	20.200	24
JUN06_97	19.683	24	20.192	24	19.242	24
JUN07_97	19.500	24	19.888	24	19.117	24
JUN08_97	20.738	24	21.071	24	20.371	24
JUN09_97	21.371	24	21.800	24	20.975	24
JUN10_97	21.183	24	21.567	24	20.779	24
JUN11_97	21.183	24	21.658	24	20.842	24
JUN12_97	21.804	24	22.229	24	21.371	24
JUN13_97	22.370	23	22.857	23	21.939	23

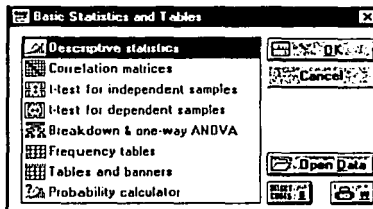
M 25. Cuadro con los resultados. Medias de las variables de temperatura

Eligiendo el contenido completo de la tabla, copiando y pegando en Word se editaron y presentaron para elaborar el Volumen de Temperatura presentado en el ANEXO C.

PROCESO: Obtención de estadística descriptiva

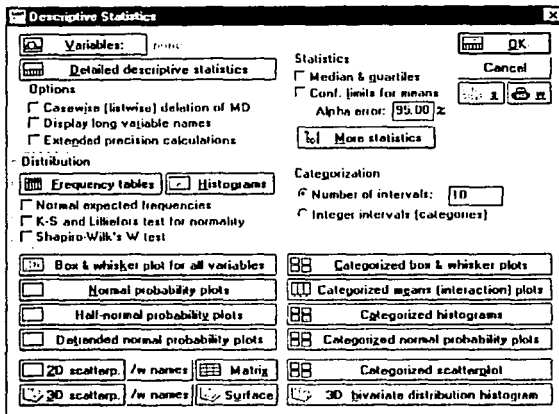
MACRO: Ctrl + F

- 1) Utilizando el módulo Basic Statistics and Tables, con la combinación de teclas Ctrl. + S, aparece el siguiente cuadro de diálogo y utilizamos la primera opción correspondiente a la estadística descriptiva



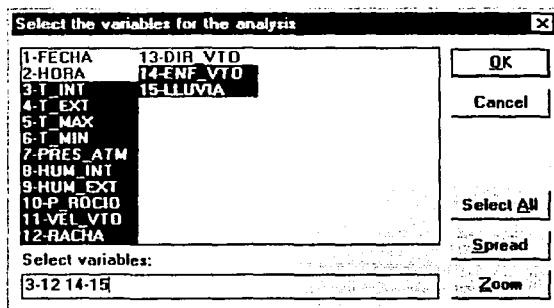
M 26. Panel inicial del módulo Basic Statistics and Tables

- 2) Al presionar el botón OK, aparece la pantalla que sigue, en donde se especifica la variable de interés.



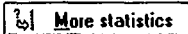
M 27. Diálogo para la estadística descriptiva

- 3) Para elegir las variables a analizar, se hace como aparece en esta imagen

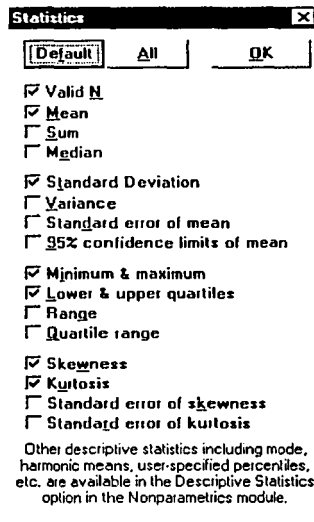


M 28. Selección de variables para reporte estadístico

4) Después de aceptar con OK, regresa al diálogo de la figura M 27, ahí se da un clic sobre el botón



y aparece la siguiente pantalla:

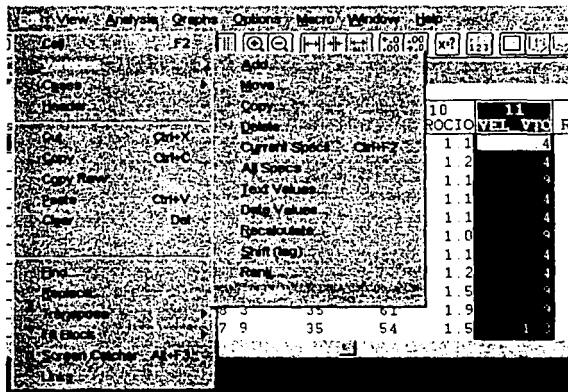


M 29. Elige el cálculo de diferentes estadísticas

A las opciones por default se agregó la que dice "Lower & upper quartiles", "Skewness" y "Kurtosis", se acepta con OK regresa al cuadro de la figura M 27, se elige nuevamente OK, y así se genera el cuadro de resultado.

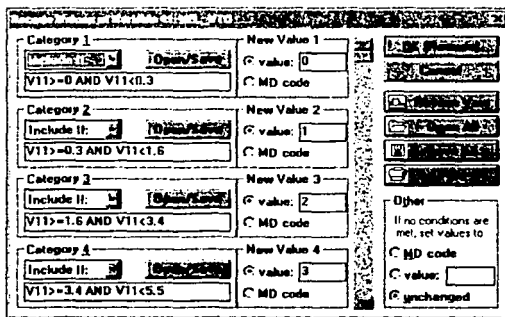
**PROCESO: Creación de tabla de contingencia (Rosa de los vientos)
MACRO: Ctrl + Z**

- 1) En el módulo de *Basic Statistics and Tables*, marcada previamente la variable VEL_VTO, escoge la opción "RECODE" de las operaciones sobre variables.



M 30. Cuadro para seleccionar función recode

- 5) Aparece el cuadro de diálogo de la siguiente figura, ahí da un clic en el botón "OPEN ALL", para tomar el archivo ESCALAB.TXT en donde se encuentran los valores que convierten las velocidades a la escala Beaufort que deseamos, como se ve aquí.



M 31. Cuadro para modificar los valores que toma la variable VEL_VTO

- Al dar OK, pregunta si se desea recodificar la variable y elige YES, asigna los valores.
- 6) Para crear propiamente la tabla de contingencias, entra al panel inicial del módulo de estadísticas básicas (M 27) y escoge la opción *TABLES AND BANNERS*, apareciendo el siguiente cuadro de diálogo.

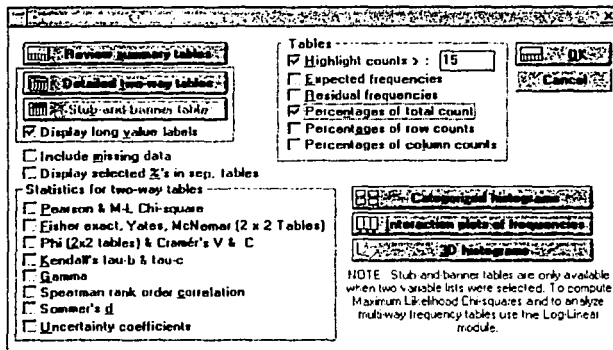
M 32. Cuadro para seleccionar el tipo de tabla de contingencia.

- 7) Elige el botón SPECIFY TABLES para seleccionar las variables que formaran la tabla, en este caso la velocidad y la dirección del viento.

M 33. Cuadro para seleccionar las variables de la tabla de contingencia.

- 8) Da OK para aceptar y entonces despliega el cuadro de diálogo que sigue, en donde se marca la opción "Percentages of total count", para saber las frecuencias relativas correspondientes.

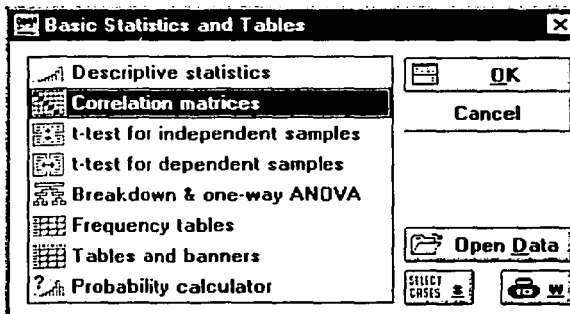
**TESIS CON
FALLA DE ORIGEN**



M 34. Selección del tipo de tabla de contingencia.

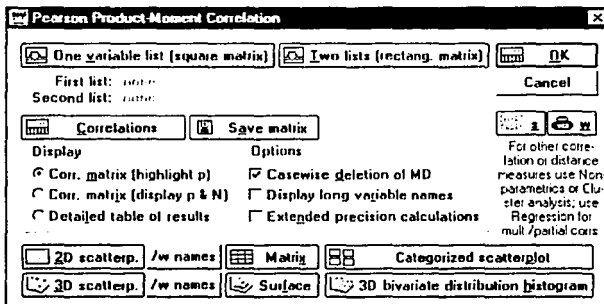
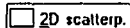
PROCESO: Gráficas de dispersión
MACRO: Ctrl + G

- 1) Elige en el módulo *Basic Statistics and Tables* la segunda opción "Correlation matrices"



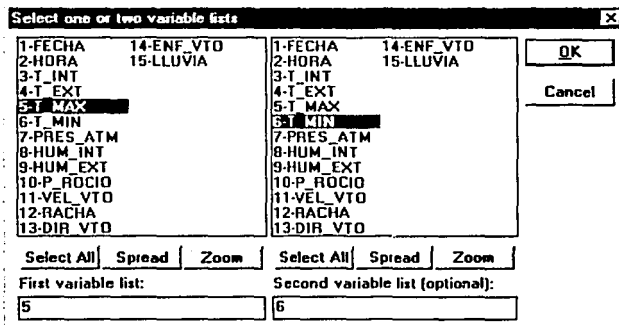
M 35. Módulo de estadísticas básicas. Matrices de correlación

- 2) Enseguida muestra el cuadro para elegir el análisis o gráfica a realizar. Para las gráficas que deseamos debe elegirse el botón:




M 36. Cuadro de diálogo para correlación y gráficas de dispersión

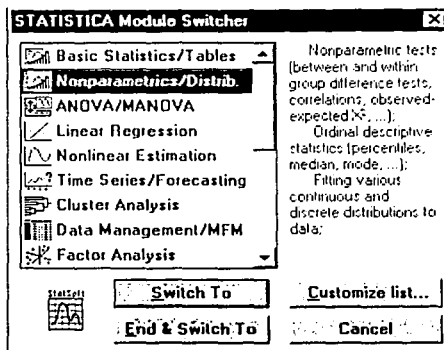
- 3) Elige las variables a utilizar en la gráfica mediante el cuadro que sigue y así se realiza la gráfica de dispersión de las variables seleccionadas.



M 37. Selección de las variables a graficar.

PROCESO: Comparación de dos medias MACRO: Ctrl + U

- 1) Con el botón  aparece el Module Switcher, elige la opción de pruebas no paramétricas y cambia a dicho módulo. Ya previamente está abierto el archivo que contiene los datos



M 38. Selección de pruebas no paramétricas

- 2) Elige la quinta opción del menú, correspondiente a la prueba U. Nótese el formato del archivo.

#	g-canal	13	CON
1	g-canal	14	900
2	g-canal	17	000
3	g-canal	18	600
4	g-canal	19	300
5	g-canal	19	100
6	g-canal	18	200
7	g-canal	18	100
8	g-canal	18	400
9	g-canal	18	100
10	g-canal	17	000
11	g-canal	15	400
12	g-canal	13	800
13	p3	14	300
14	p3	15	600
15	p3	18	900
16	p3	22	000
17	p3	22	900
18	p3	21	400
19	p3	19	300
20	p3	18	500
21	p3	18	100
22	p3	16	600
23	p3	17	400
24	p3	15	200

Nonparametric Statistics & Distribution Fitting

Nonparametric Statistics:

- 2 x 2 Tables χ^2 /Fisher, McNemar, Fisher exact
- Observed versus expected χ^2
- Correlations (Spearman, Kendall tau, gamma)
- Wald-Wolfowitz runs test
- Mann-Whitney U test
- Kolmogorov-Smirnov two-sample test
- Kruskal-Wallis ANOVA, median test
- Sign test
- Wilcoxon matched pairs test
- Friedman ANOVA & Kendall's concordance
- Cochran Q test
- Ordinal descriptive statistics (median, mode, ...)

Continuous Distrib.:

- Normal
- Rectangular
- Exponential
- Gamma
- Log normal
- Chi-square
- Others ...

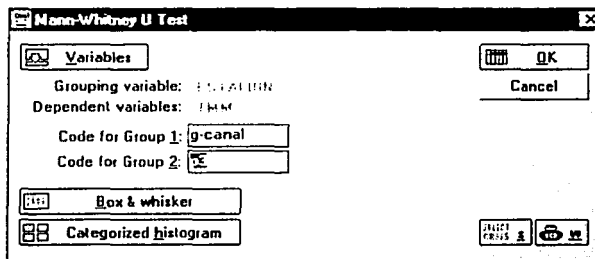
Discrete Distrib.:

- Binomial
- Poisson
- Geometric
- Bernoulli

Buttons: OK, Cancel, Open Data, Print, Help

M 39. Elige prueba U de Mann Whitney

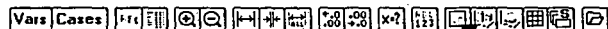
- 3) En el siguiente cuadro de diálogo al seleccionar el botón de "variables" elige como primer variable "ESTACION" y como segunda la temperatura media mensual (TMM). Y con doble clic se eligen los campos de "Code for Group1" y también en el 2. Presiona OK:



M 40. Elección de variables a analizar

GRÁFICAS EN DOS DIMENSIONES

Haciendo un clic en el nombre de la variable que deseamos graficar, se resalta el contenido de toda la columna correspondiente y



IA	2 HORA	3 T INT	5 T MAX	6 T MIN	7 PRES ATM	8 HUM INT	9 HUM EXT	
_97	1	16.3	10.6	10.8	10.3	788.5	45	71
_97	2	16.2	9.8	9.8	9.4	788.5	45	73
_97	3	16.2	8.9	9.0	8.8	788.4	45	75
_97	4	16.1	8.4	8.5	8.3	788.3	45	75
_97	5	16.0	7.7	7.5	7.1	788.4	45	81
_97	6	15.8	7.4	7.5	7.2	788.6	45	81
_97	7	15.8	6.3	6.4	6.3	788.5	45	86
_97	8	15.7	7.4	7.7	7.1	789.1	45	79
_97	9	15.7	11.5	11.0	9.9	789.3	45	65
_97	10	15.8	14.3	14.4	14.1	789.6	44	55
_97	11	15.9	15.4	15.8	14.9	789.2	44	49
_97	12	16.1	17.5	18.0	17.4	788.3	42	33

Se elige el botón correspondiente a las gráficas personalizadas en 2-D

M 41. Procedimiento para elegir una variable que se desea graficar

Se presenta el cuadro de diálogo de la siguiente figura, en el que elegimos el tipo de gráfica, nombre de variable o caso que se tomará para el eje X y el número de casos que se deben tomar para realizar la gráfica. El siguiente cuadro, y todo el procedimiento señalado hasta ahora fue utilizado para realizar la gráfica de la fig. 21.

Spreadsheet: Custom 2D Graphs [X]

Plot 1

Line Plot

X: Row Names

Y: T_EXT

Plot 2

Ignore

X:

Y:

Plot 3

Ignore

X:

Y:

Plot 4

Ignore

X:

Y:

OK

Cancel

Values From:

1


To:

432

Double-click on the field to select from list.

M 42. Cuadro para establecer los elementos de una gráfica en 2-D

Pero en las fig. 20 y 22 a 25 se deben elegir las variables y posteriormente especificar que se tome la variable HORA para ser desplegada en el eje X. Veamos el ejemplo de como se realizó la figura 20.

Después de elegir las dos variables de temperatura y el botón correspondiente , se despliega el cuadro de diálogo siguiente, en el que como se ve se especifica que los valores a graficar, se tomarán del caso 1 al 24, correspondiente a las 24 horas del primer día del mes de enero.

	4	5	6	7	8	9	10
	T_EXT	T_MAX	T_MIN	PRES_ATM	HUM_INT	HUM_EXT	P_ROCIO
21	3	12.9	13.1	12.7	788.4	33	62.5.8
21	3	12.4	12.6	12.4	700.3	22	66.5.8
21	0	12.0	12.2	11.8			
21	0	11.3	11.5	11.0			
20	9	10.4	10.7	10.2			
20	8	9.8	10.1	9.6			
20	7	9.8	9.7	9.5			
19	3	10.6	11.3	9.9			
20	3	12.8	13.6	12.0			
20	6	16.3	17.0	15.4			
21	5	19.3	19.9	18.8			

Spreadsheet: Custom 2D Graphs [X]

<p>Plot 1</p> <p><input type="button" value="Line Plot"/></p> <p>X: <input type="text" value="Row Names"/></p> <p>Y: <input type="text" value="T_INT"/></p>	<p>Plot 2</p> <p><input type="button" value="Line Plot"/></p> <p>X: <input type="text" value="Row Names"/></p> <p>Y: <input type="text" value="T_EXT"/></p>	<p><input type="button" value="OK"/></p> <p><input type="button" value="Cancel"/></p>
<p>Plot 3</p> <p><input type="button" value="Line Plot"/></p> <p>X: <input type="text"/></p> <p>Y: <input type="text"/></p>	<p>Plot 4</p> <p><input type="button" value="Line Plot"/></p> <p>X: <input type="text"/></p> <p>Y: <input type="text"/></p>	<p>Values From:</p> <p><input type="text" value="1"/></p> <p>To:</p> <p><input type="text" value="24"/></p> <p>Double-click on the field to select from list</p>

M 43. Procedimiento para hacer la gráfica de la figura 20

para cambiar el valor que tomará el eje X, se hace doble clic en el cuadro donde ahora aparece "Row Names" y nos aparece el siguiente cuadro en donde existe la posibilidad de cambiar el dato que deseamos.

Get Plotted Values From: [X]

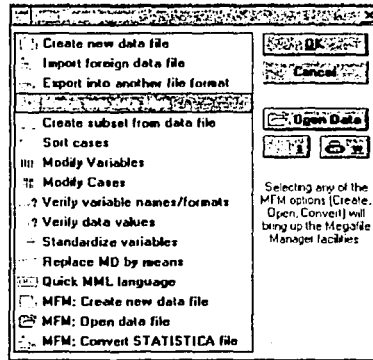
<p>FECHA</p> <p>HORA</p> <p>T_INT</p> <p>T_EXT</p> <p>T_MAX</p> <p>T_MIN</p> <p>PRES_ATM</p> <p>HUM_INT</p> <p>HUM_EXT</p> <p>P_ROCIO</p> <p>VEL_VTO</p> <p>RACHA</p>	<p><input type="button" value="OK"/></p> <p><input type="button" value="Cancel"/></p> <p>Get From</p> <p><input checked="" type="radio"/> Column</p> <p><input type="radio"/> Row</p> <p><input type="radio"/> Col Names</p> <p><input type="radio"/> Row Names</p>
---	---

M 44. Cuadro para modificar los valores que toma el eje X ó Y

Realizando el cambio para las dos variables, tendremos que en el eje X, se grafican las horas y en el eje Y las temperaturas, como se tiene en la Figura 20.

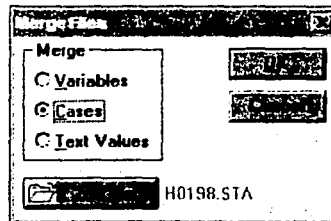
CREACIÓN DEL ARCHIVO ANUAL

- 1) Abrir el archivo del mes de enero y utilizando el módulo de manejo de datos *Data Management*, seleccionar la cuarta opción



M 45. Panel inicial del módulo para manejo de datos. Unión de archivos.

- 2) Despliega el cuadro de diálogo para unir (en este caso) CASES de el archivo abierto y un segundo archivo, el cual es elegido con el botón que dice "2nd file", por default aparece el nombre del archivo abierto en el momento por lo que debemos seleccionar el archivo de febrero y al dar OK se unen los dos archivos, salvar con otro nombre (por ejemplo 1997.sta). Este proceso se repite para cada mes, tomando como primer archivo el que deseamos contenga la información de todo el año, y como segundo archivo los meses restantes uno por uno y en el orden adecuado.

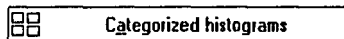


M46. Cuadro para la selección del segundo archivo y condiciones de unión

En los archivos que se crearon correspondientes a 1997 y 1998, se aumentó una variable, después de la FECHA para indicar el MES.

ELABORACIÓN DE POLÍGONOS DE FRECUENCIA ANUALES

- 1) Realizar los pasos 1 y 2 del macro para "Obtención de Temperatura media diaria". Como en el paso 3 de ese macro elegir mediante un cuadro como el de la figura M 23, la variable deseada, aceptando con el botón OK, aparece el cuadro de la figura M22, en donde hay que hacer un clic sobre el botón



- 2) Aparece una pantalla para seleccionar que los histogramas son clasificados por la variable MES (la que fue creada después de unir los archivos mensuales).

Select one or two grouping (categorization) variables			
1-FECHA	14-DIR_VTO	1-FECHA	14-DIR_VTO
2-MES	15-ENF_VTO	2-MES	15-ENF_VTO
3-HORA	16-LLUVIA	3-HORA	16-LLUVIA
4-T_INT		4-T_INT	
5-T_EXT		5-T_EXT	
6-T_MAX		6-T_MAX	
7-T_MIN		7-T_MIN	
8-PRES_ATM		8-PRES_ATM	
9-HUM_INT		9-HUM_INT	
10-HUM_EXT		10-HUM_EXT	
11-P_ROCID		11-P_ROCID	
12-VEL_VTO		12-VEL_VTO	
13-RACHA		13-RACHA	

First variable:
 Second variable:

M 47. Selección de variable para clasificar los histogramas

- 3) Posteriormente pregunta por medio del siguiente cuadro, cuáles son los meses que se van a considerar para la clasificación, elegimos el botón "ALL"

Select codes for the grouping variables: X

M 48. Cuadro para definir los códigos del mes deseado

- 4) Así crea una gráfica con histogramas para cada mes, pero para cambiar a polígonos de frecuencia hacer doble clic en cualquier esquina de la gráfica y aparece el cuadro siguiente, en donde debe elegirse el tipo de gráfica lineal como se muestra aquí. Aceptando con OK.

Graph Type:

Layered

TITLES

Title 1: Histogram: PRES_ATM:...

Histogram: PRES_ATM: PRESION ATMOSFERICA (MB)

SCALES

Max: 802 Scaling: Manual

 Step: 2 Type: Linear

 Min: 778

BACKGROUND

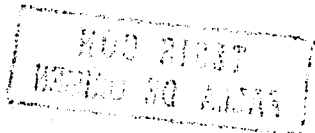
Outside:

 Inside:

M 49. Cuadro para definir el formato de las gráficas

GLOSARIO

- DDE (Dinamic Data Exchange) [Intercambio dinámico de información]. - En Windows de Microsoft, OS/2 y en el System 7 de Macintosh, canal de comunicación entre procesos (IPC, por sus siglas en inglés) a través del cuál los programas pueden intercambiar activamente información y controlar otras aplicaciones. Para que pueda existir un intercambio dinámico de información, los programas deben cumplir con las especificaciones de Microsoft Corporation.
- DLL (Dynamic Link Library) [Biblioteca de enlaces dinámicos]. - Los archivos DLL contienen por lo general rutinas que dos o más programas pueden compartir.
- Hardware. - componentes electrónicos, tarjetas, periféricos y equipo que conforman un sistema de cómputo.
- Meteoro. - (del griego meteoron) Cualquier cosa que pasa en el aire, es decir, cualquier fenómeno atmosférico como el trueno, el rayo, el granizo, la lluvia, etc.
- Meteorología. - (del griego meteoros, meteoro y logos, tratado). Ciencia que estudia los fenómenos atmosféricos, especialmente en orden a la precisión del tiempo.
- ODBC(Open Data Base Connectivity). - Es una interface de programación de Microsoft que proporciona un lenguaje común para aplicaciones de Windows, para el acceso a bases de datos en una red.
- OLE (Object Linking and Embedding) [Vinculación e inserción de objetos]. - Conjunto de normas (desarrolladas por Microsoft Corporation e incorporadas a Windows de Microsoft y al software del sistema Macintosh de Apple) que permiten crear vinculaciones dinámicas y de actualización automática entre documentos, así como insertar un documento creado en cierta aplicación en otro documento creado en otra aplicación.
- Portable [portátil]. - Capaz de trabajar en una variedad de plataformas de hardware.



- Shareware** [programas compartidos]. - Es el software que se puede probar antes de comprarlo. No es realmente un tipo de software, sino un método de distribución, pues permite probar un programa para asegurarse que es adecuado a las necesidades propias. Estos programas de computación cuentan con derechos de autor; así, si al usuario le agrada alguno y decide usarlo, se cuenta con que remitirá una cantidad al autor.
- Software.** - programas de sistema, utilerías o aplicaciones expresadas en un lenguaje legible para las computadoras.
- Teachware/Learnware** [software de enseñanza-aprendizaje]. - término empleado para los programas de cómputo creados con fines de enseñanza y aprendizaje en cualquier materia de conocimientos.

TESIS CON
FALLA DE ORIGEN