UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

MODELO DIGITAL
PARA AJUSTE DE DATOS HIDROMETRICOS A DISTRIBUCIONES TEORICAS DE PROBABILIDAD.

JESUS FLORES REYES

TESIS PROFESIONAL

México, D. F.

1983

Dedico este trabajo:

A mis padres por su estímulo y apoyo moral.

A la Universidad Nacional Autónoma de México que me brindó la oportunidad de ingresar a sus aulas y especialmente a la Facultad de Ingeniería, a quien debo mi formación profesional.

A mis profesores, de cuyas enseñanzas estoy profundamente agradecido.

A mis compañeros de aulas y amigos con quienes he tenido la fortuna de convivir.

Jesús Flores Reyes.

FACULTAD DE INGENIERIA
EXAMENES PROFESIONALES
60-I-34

UNIVERSIDAD NACIONAL
AUTONOMA

Al Pasante señor JESUS FLORES REYES,
P r e s e n t e .

En atención a su solicitud relativa, me es grato transcribir a usted a continuación el tema que aprobado por esta Dirección propuso el Profesor Ing. Roberto Carvajal Rodríguez, para que lo desarrolle como tesis en su Examen Profesional de Ingeniero CIVIL.

"MODELO DIGITAL PARA AJUSTE DE DATOS HIDROMETRICOS A DISTRIBUCIONES TEORICAS DE PROBABILIDAD"

1. Introducción.
2. Identificación del problema y alternativas de solución.
3. Determinación de la metodología adecuada.
4. Estructuración, desarrollo y operación del modelo.
5. Ejemplo de aplicación.
6. Conclusiones.

Ruego a usted se sirva tomar debida nota de que en cumplimiento de lo especificado por la Ley de Profesiones, deberá prestar Servicio Social durante un tiempo mínimo de seis meses como requisito indispensable para sustentar Examen Profesional; así como de la disposición de la Dirección General de Servicios Escolares en el sentido de que se imprima en lugar visible de los ejemplares de la tesis, el título del trabajo realizado.

A t e n t a m e n t e
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria, 20 de febrero de 1981
EL DIRECTOR

ING. JAVIER JIMENEZ ESPRIU

JJE/RCLR/ser

I N D I C E

Glosario.

v.a.	variable aleatoria
f.d.p.	función densidad de probabilidad
F.D.P.	función de distribución de probabilidad
D.P.A.	distribución de probabilidad acumulada
$\bar{x}$	media aritmética muestral
s	desviación standard muestral
μ	media aritmética poblacional
σ^2	varianza poblacional
σ	desviación standard poblacional
C.v.	coeficiente de variación.
α_3	coeficiente de sesgo
α_4	coeficiente de curtosis
α, β, γ	parámetros
ν	grados de libertad
F	función

I.- INTRODUCCION.

La sociedad genera cada vez más problemas de magnitud creciente en forma exponencial, sean éstos: de carácter social, polîtico ó económico; lo que hace importante destacar la trascendencia de encontrar las soluciones relativas, mismas que deben ser adecuadas, factibles y de costo óptimo; pués misión fundamental de las disciplinas del conocimiento, es plantear alternativas de solución al estado de conflicto en que la sociedad se encuentra.

De lo anterior cabe observar, que la ingeniería por su propia naturaleza se encuentra en una situación inmejorable para satisfacer en la medida de su aplicabilidad, un alto porcentaje de los problemas generados.

El objetivo del presente trabajo, es plantear una solución a la interrogante de conocer las características intrînsecas en una muestra de variables aleatorias y estar en oportunidad de inferir información de la población relativa; con una confiabilidad aceptable y un ahorro de tiempo. Este modelo digital, liga conceptos de probabilidad, estadística y computación; la aplicación adecuada del mismo, determina con un grado de confianza el comportamiento de algunos fenómenos fîsicos.

II.- IDENTIFICACION DEL PROBLEMA Y ALTERNATIVAS DE SOLUCION.

II.1.- Planteamiento.

A lo largo de su historia, el hombre ha desarrollado la ciencia como parte de un esfuerzo orientando a comprender el mundo y mejorar de esta forma sus condiciones de vida; para ello, observa los fenómenos que -- ocurren a su alrededor y posteriormente experimenta con ellos, recogiendo el resultado de sus observaciones mediante el uso del lenguaje. De esta manera, elabora una descripción del fenómeno en lenguaje ordinario, lo que constituye de hecho un modelo de dicho fenómeno. Se -- puede utilizar más de un lenguaje para describir el mundo físico, algunos con ventajas sobre otros; tal es el caso de las matemáticas que aportan entre otras ventajas la de su precisión. A objeto de pasar de la concepción meramente fenomenológica en lenguaje ordinario a - la concepción matemática del fenómeno, debe existir una traducción que permita asociar un término y una proposición dentro de la matemática, a cada término y proposición de la descripción original; misma que debe tomar - en cuenta la sintaxis de ambos lenguajes para obtener una interpretación aceptable.

Es frecuente que en la ingeniería y disciplinas afines, se trabaje con variables aleatorias de un espacio muestral, mismo que en la mayoría de los casos, no es representativo de la correspondiente población y la información que de él se infiera respecto a la misma, - es por consecuencia poco confiable; lo que dificulta la interpretación aceptable del fenómeno en estudio, así - como la aplicación de las técnicas existentes encaminadas a su respectiva solución.

La teoría básica empleada en el presente trabajo no es nueva; sin embargo, ensamblar las diversas componentes y darles una secuencia en su ejecución, requiere seguir un criterio delineado con anterioridad; para lo cual, el planteamiento general obedece a la lógica tradicional,útil al resolver problemas de ingeniería.

II.2.- Fundamentos Teóricos y Prácticos válidos para la creación del modelo.

Es rigurosamente cierto que al enfrentar de una ú otra forma problemas que involucran variables aleatorias, el primer obstáculo es el de recopilar la información respectiva, misma que una vez conformada, en un -- buen número de casos no es suficientemente confiable. - Sin embargo, por las razones que se argumenten puede -- creerse que este primer escollo saludablemente se deja a un lado y así, como consecuencia propia de la búsqueda de soluciones, surgen una serie de interrogantes relativas a la naturaleza, tamaño y tipo de la muestra -- más apropiada, etc.. El análisis adecuado en esta etapa de la solución es el de muestreo estadístico, de tal - suerte, que al probar con ajuste a distribuciones de -- probabilidad para diferentes estadísticas muestrales, exista una correspondencia entre una parte y otra, lo - que puede conseguirse mediante iteraciones; llendo de - lo primero a lo segundo y viceversa. Cabe aclarar, que el presente modelo sólo es útil en lo relativo a la segunda parte mencionada.

Básicamente, la estadística puede ser: descriptiva o deductiva e inferencial o inductiva. Cuando una -- muestra de variables aleatorias es representativa de -- una población, se puede inferir información acerca de - ésta, mediante un cuidadoso análisis de aquella; lo que conduce a trabajar en el ámbito de la estadística inferencial. Pero tales inferencias no son absolutamente --

ciertas, lo que hace necesario conocer la probabilidad correspondiente a tales deducciones; luego entonces, la estadística descriptiva será solamente el primer paso en la solución del problema.

Una vez analizados los datos muestrales, se estará en mejor oportunidad de decidir al seleccionar la metodología adecuada a un problema determinado; como podría ser el caso de simulación hidrológica, problemas que involucren líneas de espera o proyección financiera; así mismo en lo que respecta al cálculo de parámetros estadísticos, éste será más confiable, puesto que se habrán calculado con las expresiones correctas.

El contar con un modelo digital en este tipo de problemas es de gran ayuda, pues si el usuario está familiarizado con el tema, le bastará correr el programa é interpretar y análizar la información resultante, lo que traerá un ahorro considerable de tiempo, siempre y cuando, como ya se dijo, exista una técnica de muestreo implementada y aceptada para el problema específico.

II.3.- Alternativas de solución.

Existen un buen número de técnicas para resolver problemas de estas características, cada cual tiene sus propias ventajas y limitaciones.

Una alternativa frecuentemente usada, es considerar que al relacionar dos o más de las variables implicadas en el problema que se trate, se estará en oportunidad de determinar una expresión analítica consistente en tener una variable como independiente, función de otra (s), según que se trate de una regresión simple ó múltiple respectivamente. Sin embargo, esto será una parte del problema y tal vez la más fácil, pero en realidad la importancia de este análisis de regresión, estriba en conocer el grado de confianza que deba tenerse al respecto; por lo que el complemento de esta tarea es utilizar una analisis de correlación, con el o los coeficientes respectivos, determinando niveles de confianza lo que hace casi imperioso conocer la (s) distribución (s) de probabilidad, a la que se ajusta las variables en cuestión; lo que permite comparar un resultado- con otro, y si se quiere, generar entonces una muestra- sintética.

Cabe aclarar respecto a lo anterior, que la expresión - de regresión puede ser la de una recta, un polinomio, - una cónica, etc., dependiendo este del propio fenómeno- y de la experiencia que sobre el mismo tenga el análista.

Otro criterio útil es el de utilizar series de tiempo - que en problemas específicos redituan beneficios, pues- parte de un conjunto de observaciones hechas en instantes determinados, normalmente a intervalos iguales de - tiempo, lo que lleva a una expresión de la forma :

$$y = f (t)$$

donde: t = tiempo

En la figura. (1) pueden observarae variaciones características que se presentan en diferentes grados; tales variaciones o movimientos pueden clasificarse en cuatro tiempos principales:

Movimientos seculares.- que se refiere a la dirección - general que la serie de tiempo parece seguir y cuya curva de tendencia puede ajustarse por mínimos cuadrados.

Movimientos Cíclicos.- referentes a las osclaciones de- larga duración alrededor de la curva de tendencia-----

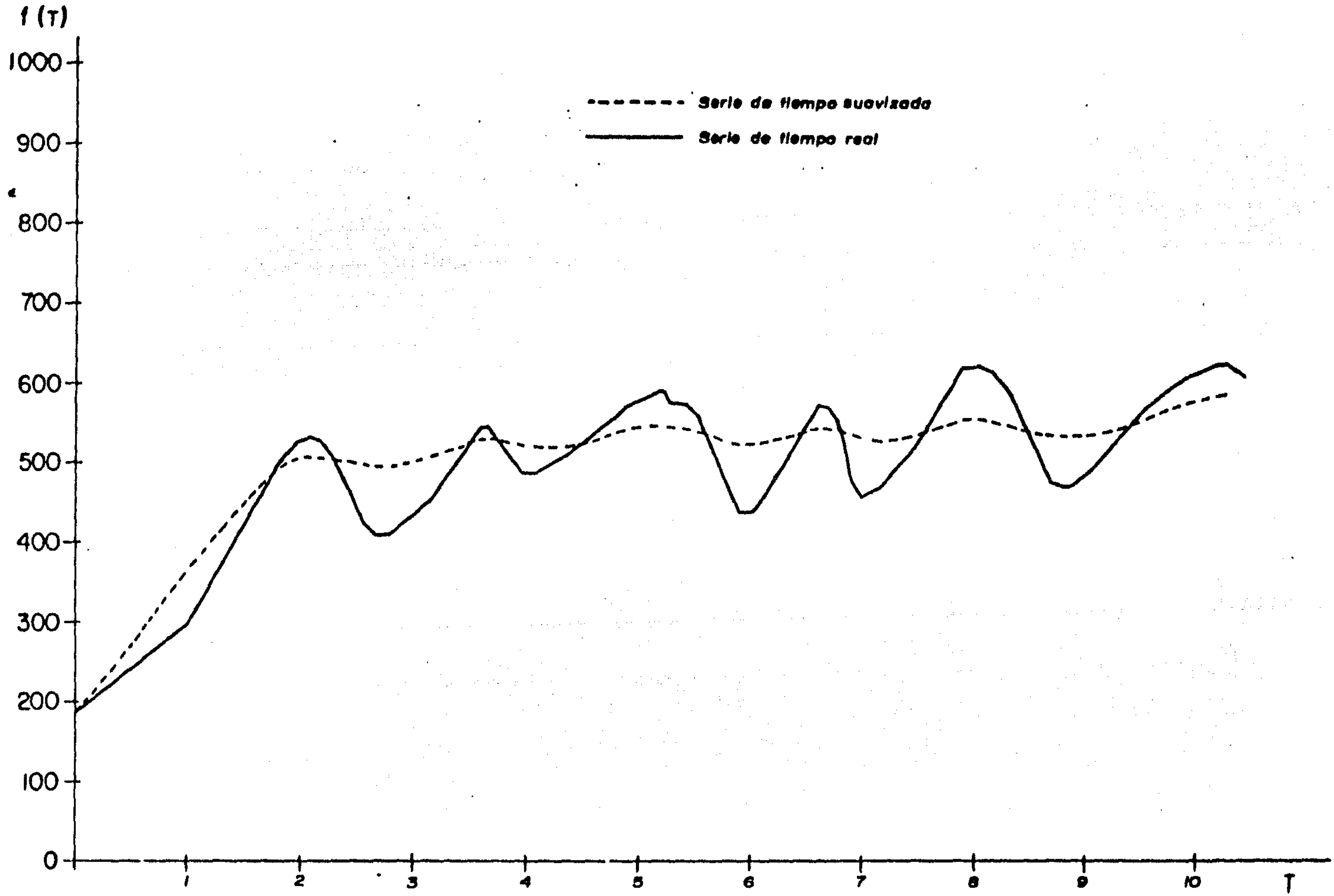


Fig. 1. Gráfica de una serie de tiempo.

mismos que pueden ser periódicos ó no.

. Movimientos Estacionales.- relativos a las normas casi idénticas que una serie de tiempo parece seguir durante los correspondientes meses de los sucesivos años.

. Moviemientos al Azar.- considerados como esporádicos, debidos a sucesos ocasionales, que duran un corto tiempo y tan intensos que pueden originar un nuevo ciclo.

Cabe observar, que una serie real y una ideal, cuando más difieren es cuando se presenta un movimiento aleatorio. Si T, C, S é I, representan respectivamente, los cuatro casos anteriores, el resultado sería el producto*:

$$Y = TxCxSxI$$

Como consecuencia de lo anterior, puede decirse que el análisis de series de tiempo consiste básicamente en una investigación de sus componentes.

Puntos importantes en el análisis de series de tiempo:

1.- Colección de datos. (teniendo presente el propósito que se persigue).

2.- Representar la serie de tiempo, anotando cualitativamente la presencia de tendencia de larga duración, variaciones cíclicas y estacionales.

3.- Construir las curvas de tendencia de larga duración y obtener los valores de tendencia (T).

4.- De ser el caso, obtener un índice estacional y desestacionalizar los datos.

5.- Ajustar los datos anteriores a la tendencia, y teóricamente, los resultados tendrán variaciones cí-- clicas e irregulares.

6.- Representar las variaciones cíclicas, anotan do su periocidad respectiva.

7.- Combinando los resultados de los pasos anteriores, y discutir el resultado final.

Pudiera también suponerse en forma ideal una variación senoidal para un caso particular, pero esto es más bien producto de la experiencia.

El criterio aquí expuesto, es el de ajuste a distribuciones teóricas de probabilidad que ya anterior-- mente se comentó; y cabe aclarar al respecto, que dada la incertidumbre con aportes de pánico que frecuente-- mente se presenta en los comentarios sobre estadística inferencial y probabilidad, la modesta idea plasmada en este trabajo, no es inovación y sí; un intento por cono cer un poco más a detalle los fenómenos hidrológicos; - sin que se pretenda establecer, que estas ideas son -- las únicas para resolver el problema.

* algunos emplean $Y = T + C + S + I$

III.- DETERMINACION DE LA METODOLOGIA ADECUADA.

III.1.- Elementos y Análisis.

Cualquier cantidad X es una variable aleatoria, si para un número real x, existe la probabilidad de que X sea menor o igual que x; y X puede ser ó no, producto de observaciones. Esto es, si la cantidad X satisface la condición:

$$0 \leq P (X \leq x) \leq 1 \qquad \text{III.1.}$$

entonces, es una variable aleatoria.

Las variables aleatorias, pueden ser discretas ó contínuas; para el primer caso, la v.a. toma un número finito ó infinito contable de valores, mientras que para el segundo caso, la v.a. toma un número infinito no contable de valores.

Sea X una v.a. contínua en el intervalo $a \leq X \leq b$, ó sea que puede tomar cualquier valor real en ese intervalo. Existe un número infinito de valores que puede adquirir la variable en el intervalo; por lo que no es posible definir su distribución de valores en forma tabular, o lo que es lo mismo, no es posible expresar explícitamente el conjunto infinito de parejas ordenadas $\{(Xi,Pi)\}$, que se tiene para todo valor de X en el intervalo. En este caso, deberá definirse la f.d.p., de X en forma gráfica o analítica.

Por la continuidad de los valores de X en el intervalo, en este caso el polígono de probabilidad se convierte en una curva contínua de ecuación $y = f(x)$, siendo f(x) una función real de la variable aleatoria X en su intervalo de definición, ó f(x) será la f.d.p. de la v.a. X, y su expresión algebraica constituye la representación analítica de la f.d.p. de la variable. Esta expresión analítica se determina de manera que la probabilidad $P (X_1 \leq X \leq X_2)$, en donde el intervalo $X_1 \leq X \leq X_2$, está completamente contenido en el intervalo $a \leq X \leq b$ de definición de la v.a. X; y es igual al área bajo la curva de la fun---

ción f(x), entre $X = X_1$ y $X = X_2$. ó

$$P\ (X_1 \leq X \leq X_2) = \int_{x_1}^{x_2} f(x)\ dx \qquad \text{III.2.}$$

La función f(x) debe cumplir con las condiciones siguientes para que sea f.d.p., de una v.a. contínua:

i.- f(x) es no negativa para todo valor de X.

$$f(x) \geq 0 \qquad \text{ó} \qquad a \leq X \leq b$$

ii.- El área bajo la curva de la función f(x) en el intervalo $a \leq x \leq b$, vale uno.

$$\int_a^b f(x)\ dx = 1 \qquad \text{III.3.}$$

Fundamentalmente puede definirse la D.P.A., de una v.a. contínua X, por medio de su f.d.p., que establezca la relación entre los valores de la variable - en un intervalo y la probabilidad de que ella adquiera esos valores.

La D..P.A., se puede representar en forma tabular, gráfica o analítica:

i.- En forma tabular por medio de la escritura explícita de todos los valores de la v.a., y sus correspondientes probabilidades.

ii.- En forma gráfica a través del conjunto - de puntos $\{(X_i, P_i)\}$ dibujados en un - sistema de ejes coordenados, uniendo los puntos sucesivos por segmentos de recta.

iii.- En forma analítica a través de una función real P(x), tal que el valor de la - función, cuando la v.a. X toma el valor X_i, es igual a la probabilidad de que --

ocurra este valor, o sea, de que ocurra el evento E_i.

La D.P.A. puede definirse formalmente por medio de una función:

$$f(x_c) = P(X \leq x_c) \qquad \text{III.4.}$$

donde el valor de esta función es la probabilidad del evento en donde la v.a., toma cualquier valor igual ó menor a x_c.

i.- Caso discreto.

Si la v.a., es discreta, definida en .... $a \leq x \leq b$, entonces:

$$F(x_c) = P(x \leq x_c) = \sum_{x_i=a}^{x} P(x_i) \qquad \text{III.5.}$$

ii.-Caso contínuo.

Si la v.a., contínua definida en $a \leq X \leq b$ y $f(c)$ es una f.d.p., la función $F(x)$ es:

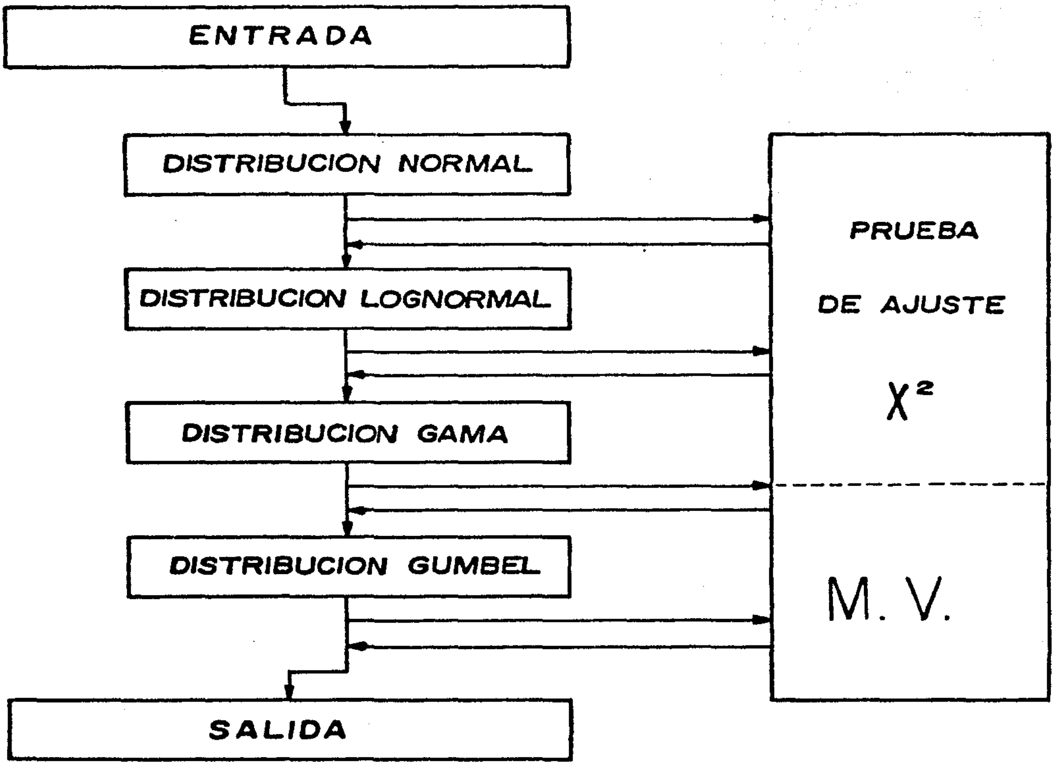
$$F(x_c) = p(X \leq x_c) = P(a \leq X \leq x_c) = \int_a^{x_c} f(x)\,dx \qquad \text{III.6.}$$

y en estos casos, discreto y contínuo $F(x)$ es una función monotónicamente creciente, así mismo, en ambos se debe cumplir que:

$$\left.\begin{array}{l} i.-\ f(x) \geq 0 \\ \\ ii.\ \sum_{x} f(x) = 1 \end{array}\right\} \qquad \text{III.7.}$$

Debe observarse la diferencia entre distribución empírica de v.a., y distribución teórica de v. a., --- puesto que la primera es producto de observaciones o experimentos registrados y la segunda, corresponde a una expresión analítica que es la relación entre estos.

En hidrología se cuenta con v. a. discretas o contínuas, pero esto no es obstáculo en el presente desarrollo, puesto que si es el primer caso, bastará incorporar una variación contínua al espacio muestral- y viceversa si se tiene al segundo y se necesitará - lo primero.



- χ^2 JI CUADRADA
- M.V. MAXIMA VEROSIMILITUD

Fig. 2 ESTRUCTURA BASICA DEL PROGRAMA TEST.

PARAMETROS DESCRIPTIVOS DE LA DISTRIBUCION

DE TENDENCIA CENTRAL III.8

MEDIA

$$\bar{X} = \sum_{i=1}^{k} \frac{X_i f_i \Delta X_i}{n}$$

Xi = Es el valor representativo de un intervalo de clase

Fi = Frecuencia correspondiente a Δi

k = Número total de intervalos de clase.

n = Número de datos de la muestra.

MEDIANA.

$$M = L_1 + C \left[\frac{\frac{n}{2} - (\sum f)_1}{fm} \right]$$

M= Aquel valor tal que la mitad de las observaciones son menores que ese valor y la otra mitad mayores que el mismo. Para "n" impar corresponde al elemento situado en la posición (n+1) /2 y para "n" par se toma el promedio de los dos centrales.

L_1 = Limite inferior o intervalo para el que fi mayor que 0.5..

c = Amplitud del intervalo.

$(\sum f_i)$ = Suma de frecuencias de los intervalos anteriores aquel en donde esta alojada la mediana.

fm= Frecuencia del mismo intervalo de clase.

MODA:

M'= Observación que se presenta con mayor frecuencia.

$$M' = L_i + C \frac{d_1}{d_1 + d_2}$$

L_i = Limite inferior del intervalo de clase con mayor frecuencia.

C = Amplitud del intervalo modal.

d_1 = Diferencia en valor absoluto de la frecuencia del intervalo modal y la frecuencia del intervalo de clase anterior al modal.

d_2 = Idem al anterior solo que con el intervalo de clase que sigue al modal.

Cuando la distribución es moderadamente asimétrica.

$$M' = \bar{X} - 3 (\bar{x} - M)$$

R A N G O

$RN = X max - X min$

DE DISPERSION III.9

DESVIACION

$$S_x = \left[\sum_{i=1}^{k} \frac{(X_i - X)^2 \; f_i \, \Delta X_i}{n - 1} \right]^{1/2}$$

Donde las variables con excepción de S_x tienen el mismo significado que en la descripción de la media.

VARIANCIA.

La variancia se denota por Sx^2

COEFICIENTE DE VARIACION

$$C.v. = \frac{Sx}{\bar{X}}$$

COEFICIENTE DE ASIMETRIA

$$C.AS. = \frac{\bar{X} - M}{Sx}$$

Vale cero para distribuciones simétricas y otro valor para asimétricas.

DE APLANAMIENTO III.10.

CURTOSIS

$$\alpha 4 = \frac{M4}{M_2^2}$$

M4 = Cuarto momento con respecto a la media é igual a $E\left[(X - \bar{x})\right]^4$

M_2^2 = Segundo momento con respecto a la media.

Debe observarse que $\alpha 4$ vale tres en una distribución mesocúrtica como la normal, mayor de tres en leptocúrticas (más alta y estrecha que la normal) y menor de tres en platicúrticas (más ancha y baja que la normal).

OTROS ELEMENTOS UTILES III.10'

ESPERANZA MATEMATICA.

$$E(x) = X_j, P(X=X_i) + \ldots + X_n P(X=X_n)$$

$$E(x) = \sum_{j=1}^{n} X_j f(X_j)$$

Cuando las probabilidades son todas iguales.

$$E(x) = \frac{X_1+X_2 + \ldots X_n}{n}$$

Xi= Variable aleatoria.

VARIANZA

$$VAR\,(X) = E\left[(X - \bar{x})^2\right]$$

$$VAR\,(X) = \left[(X - \bar{x})^2\right] f\,(X)$$

Cuando las probabilidades són todas iguales.

$$S_x^2 = \left[(X_1 - \bar{x})^2 + (X_2 - \bar{x})^2 + \ldots + (X_n - \bar{x})^2\right] / n$$

VARIABLE ALEATORIA NORMALIZADA

$$X^* = \frac{X - \bar{X}}{Sx}$$

X^* = Variable aleatoria normalizada con $\bar{X} = 0$ y $Sx = 1$, y tiene como propiedad adyacente el ser una cantidad - adimensional.

* Las variables normalizadas son útiles para comparar- diferentes distribuciones

COEFICIENTE DE CORRELACION

$$P = \frac{Sxy}{SxS_y}$$

Sxy= covarianza de (x,y)

como $|\sigma_{xy}| \leq \sigma_x \sigma_y$ implica que

$$-1 \leq p \leq 1$$

desigualdad de CHEBYSHEV

$$P\left(|x-\mu| \geq \varepsilon\right) \leq \frac{S_x^2}{\varepsilon^2}$$

si $\varepsilon = K\sigma_x$

$$P\left(|x-\mu| \geq K\sigma_x\right) \leq \frac{1}{K^2}$$

este teorema revela una propiedad general de las v.a. discretas con $\bar{X}$ y $S_{\bar{x}}^2$ finitas. En la ecuación anterior ε es cualquier número positivo y quiere decir literalmente que la probabilidad de que X difiera de su media en más de K desviaciones es menor que o igual a $1/K^2$. Equivalente, es decir que señala la probabilidad de que X se encuentra dentro de las K desviaciones cuando todavía no se ha determinado la distribución de probabilidad de X.

LEY DE LOS GRANDES NÚMEROS (EN FORMA DEBIL)

$$\lim_{n \to \infty} P\left(\left| \frac{Sn}{n} - \bar{\mu} \right| \geq \varepsilon \right) = 0$$

Este teorema establece que la probabilidad de que la media aritmética Sn/n difiera de su valor esperado por mas de ε, tiende a cero cuando $n \to \infty$

$$Sn = X_1, + X2 + \ldots Xn$$

Y X_1, X2...Xn son variables aleatorias mutuamente - independientes cada una con media μ varianza σ^2 finitas.

Serie de Taylor:

Una serie de la forma:

$$A_0 + A_1 (x-b) + A_2 (x-b)^2 + \dots , \qquad \text{III.11.}$$

o

$$\sum_{n=0}^{\infty} A_n (x-b)^n \qquad \text{III.12.}$$

en donde A y b son números reales, se llama serie de potencias de x-b.

Es evidente que la serie de potencias III.11., converge a A_0 cuando $x = b$.

Si en III.12., el número b se toma igual a cero, entonces queda:

$$\sum_{n=0}^{\infty} A_n X^n \qquad \text{III.13.}$$

la cual se llama serie de potencias de X.

Para estudiar algunas propiedades de las series de potencias, es suficiente utilizar la forma III.13., y cualquier resultado así enunciado puede también aplicarse a la serie III.12.

Teorema 1.- Si la serie de potencias III.12., es convergente en X_0, donde $X_0 \neq 0$, la serie es absolutamente convergente en cualquier número X_1, para el cual $|X_1| < |X_0|$.

Teorema 2.- Si la serie III.12., es divergente en X_1, entonces es divergente en cualquier número X_0 tal que . . . . . $|X_0| > |X_1|$.

Teorema 3.- Supóngase que la serie de potencias III.12. tiene el intervalo de convergencia R.

Sea F la función definida por III.12.,

$$F(x) = \sum_{n=0}^{\infty} An \ (x-b)^n, \qquad x \in R.,$$

luego entonces:

i.- La función F es contínua y derivable sobre x.

ii.- La serie:

$$\sum_{n=0}^{\infty} DxAn \ (x-b)^n = \sum_{n=0}^{\infty} n \ An \ (x-b)^{n-1} \qquad \text{III.14.}$$

obtenida al derivar III.12., término a término, tiene el intervalo de convergencia R, y:

$$\sum_{n=1}^{\infty} NAn \ (x-b)^{n-1} = F'(x), \ X \in R.$$

iii.-La serie:

$$\sum_{i=0}^{\infty} \left(\int_b^x An \ (t-b)^n \ dt \right) = \sum_{n=0}^{\infty} \frac{An}{n+1} \ (x-b)^{n+1}$$

obtenida al integrar III.12., término a término, tiene el intervalo de convergencia R, y:

$$\sum_{n=0}^{\infty} \frac{An}{n+1} \ (x-b)^{n+1} = \int_b^x F(t)dt; \ x \in R. \qquad \text{III.15.}$$

Teorema 4.- Considérese III.12., en la cual el término general es $An \quad (x-b)^n$.

I.- Si $\lim_{n\to\infty} \left|\frac{An+1}{An}\right| = k$, entonces:

i.- Si $k=0$, el intervalo de convergencia es $k e$.

ii.- Si $k\neq 0$, el intervalo de convergencia es:

$\left\{x \mid |x-b| < \frac{1}{k}\right\}$ es decir;

la serie es absolutamente convergente sobre la - expresión inmediata anterior y divergente sobre:

$\left\{x \mid |x-b| > \frac{1}{k}\right\}$

II.- Si $\left|\frac{An+1}{An}\right| \longrightarrow +\infty$ cuando $n \longrightarrow +\infty$

la serie es convergente únicamente en b.

Si se hace:

$$Sn(x) = \sum_{n=0}^{i} \frac{F^{(n)}(b)}{n!} (x-b)^n , \quad x \in C; \qquad \text{III.16.}$$

de manera que $\{Sn\ (x)\}$ es la sucesión de sumas parciales de la serie III.12., entonces:

$$F(x) = Sn(x) + R_{n+1}(x); \ X \in C ; \qquad \text{III.17.}$$

en donde R_{n+1} (x) es el residuo después de n+1 - términos y está dado por la fórmula:

$$R_{n+1}(X) = \frac{F^{n+1}(Z)}{(n+1)!} (x-b)^{n+1} \quad x\neq b ; \qquad \text{III.18.}$$

siendo Z un número para el cual

$b < Z < x$ si $x > b$

ó

$x < Z < b$ si $x < b$

De III.17., se observa que la sucesión de sumas parciales $\{Sn(x)\}$ converge a $F(x)$ sobre C, sí y sólo si:

$$\lim_{n \to \infty} R_{n+1} \quad (x) = 0 \quad x \in C \qquad \text{III.19.}$$

de donde se tiene el siguiente:

Teorema 5.- Sea F una función con derivada de todos los órdenes sobre un intervalo C que contiene al número b. Entonces una condición necesaria y suficiente para que la serie de Taylor para F en el punto b, represente a F sobre el intervalo C, es que se cumple III.19.

Consecuencia de lo anterior, es el siguiente Corolario:

"Para demostrar que una serie de Taylor realmente representa a la función F; no solamente debe demostrarse que la serie es convergente sobre C, sino que también debe demostrarse III.19.

Debido al comportamiento aceptable de las series de potencias, dentro de sus intervalos se pretende aquí determinar la serie de potencias que converja a $F(x) = e^x$, sobre un intervalo. La relación entre $F(x)$ y la serie es de la forma:

$$F = \left\{ (x,y) \; / \; y = \sum_{n=0}^{\infty} An(x-b)^n \; , \; x \in C \right\}; \quad \text{III.20.}$$

es decir:

$$F(x) = \sum_{n=0}^{\infty} An(x-b)^n \; , \; x \in C \; ; \quad \text{III.20'.}$$

en donde C es un subconjunto de convergencia de la serie; entonces decimos que la serie representa a la función F sobre el intervalo C.

Expresando III.... en form xtensiva:

$$F(x) = A_0 + A_1(x-b) + A_2(x-b)^2 + \dots, \; x \in C \; ; \qquad \text{III.21.}$$

y por el uso repetido del teorema 3, se concluye que, F debe poseer derivadas de todos los órdenes sobre C y por lo tanto:

$$F'(x) = A_1 + 2A_2(x-b) + 3A_3(x-b)^2 + 4A_4(x-b)^3 + \dots,$$

$$F''(x) = 2A_2 + 2.3A_3(x-b) + 3.4A_4(x-b)^2 + \dots,$$

$$F'''(x) = 2.3A_3(x-b) + 2.3.4A_4(x-b) + \dots, \qquad \text{III.22.}$$

$$\vdots$$

$$F^k(x) = (2.3.4\dots k)Ak + [2.3.4\dots k+1]\,Ak+1(x-b) + \dots,$$

Substituyendo $\underline{b}$ en lugar de x y despejando $\underline{a}$ en las igualdades anteriores, resulta que para representar a la función F mediante la serie III.12., sobre cualquier intervalo que contenga a $\underline{b}$, es necesario que:

$$A = F'(b)\;,\; A_2 = \frac{F''(b)}{2}\;,\; A_3 = \frac{F'''(b)}{2.3.}\;, \dots,\; Ak = \frac{F^k(b)}{2.3.4\dots k}$$

y la serie toma la forma:

$$\sum_{n=0}^{\infty} \frac{F^n(b)}{n!}(x-b)^n = F(b) + F'(b)(x-b) + \frac{F''(b)(x-b)}{2!} + \dots + \frac{F^n(b)(x-b)^n}{n!} + \dots,$$

III.23.

que recibe el nombre de serie de Taylor de F, en el punto b.

Para el caso específico de e^x se tiene:

$$F(x) = e^x \qquad F(0) = 1$$

$$F'(x) = e^x \qquad F'(0) = 1$$

$$F''(x) = e^x \qquad F''(0) = 1$$

$$\vdots$$

$$F^k(x) = e^x \qquad F^k(0) = 1$$

Sustituyendo estos resultados en III.23., se tiene:

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} \qquad \text{III.24.}$$

ó en forma extensiva:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \ldots, \qquad \text{III.25.}$$

Para demostrar que la serie $\sum_{i=0}^{\infty} \frac{x^i}{i!}$ converge a e^x sobre Re debe probarse que $\lim_{n \to \infty} Rn+1(x) = 0$, en donde:

$$Rn+1(x) = \frac{F^{(n+1)}(Z)}{(n+1)!} x^{n+1} = \frac{e^z x^{n+1}}{(n+1)!}$$

con $0 < z < x$ si $x > 0$

ó

$x < z < 0$ si $x < 0$.

Sí $0 < z < x$, entonces $e^z < e^x$ y:

$$\left| R_{n+1}(x) \right| < e^x \frac{x^{n+1}}{(n+1)!}; \quad x > 0$$

Si $x < z < 0$, entonces $e^z < 1$ y:

$$\left|R_{n+1}(x)\right| < \left|\frac{x^{n+1}}{(n+1)!}\right| \qquad x < 0.$$

Observe que la serie $\sum_{i=0}^{\infty} \frac{x^i}{i!}$, cuyo término general es $Gn(x) = \frac{x^n}{n!}$ es convergente sobre Re, de manera que el $\lim_{n\to\infty} \frac{x^n}{n!} = 0$ para $x \in Re$. Por tanto . . . . .

$\lim_{n\to\infty}\left|\frac{x^{n+1}}{(n+1)!}\right| = 0$ para $x \in Re$, y tenemos:

$$\lim_{n\to\infty} \left|R_{n+1}(x)\right| < e^x \lim_{n\to\infty} \frac{x^{n+1}}{(n+1)!} = 0, \; x > 0.$$

$$\lim_{n\to\infty} \left|R_{n+1}(x)\right| \quad \lim_{n\to\infty}\left|\frac{x^{(n+1)}}{(n+1)}\right| = 0, \quad x < 0$$

lo q.q.d.

III.2.- Distribuciones de probabilidad

Una F.D.P., es la más concisa representación de una distribución empírica de frecuencia; sin embargo, la selección de una de ellas no representará exactamente a la distribución observada, la bondad del ajuste en este propósito, depende entre otras cosas del tamaño de la muestra, la propia selección de la función y la mejor manera de estimar los parámetros respectivos.

Las F.D.P. pueden ser desde el punto de vista estadístico: teóricas, semi-teóricas, empíricas ó las tres. Si es deducida por axiomas y de derivaciones lógicas de la teoría de la probabilidad, la función es teórica. Si las condiciones particulares de una v.a. son acordes a una función específica de distribución de probabilidad, ó si presentan índices de que la distribución de esa v.a. pertenece a la familia de la F.D.P.; pero él ó los parámetros deben estimarse para la muestra de datos, esa -- función puede considerarse semi-teórica. Si una función es simplemente ajustada a una curva de frecuencias empíricas, la función es empírica; la misma función puede ser teória ó semi-teórica.

La selección de la F.D.P. adecuada a la distribución - empírica de frecuencias, la mejor estimación de sus parámetros, la más confiable prueba de ajuste y el consiguiente orden en -- los límites de confianza; son aspectos importantes en la aplicación de la estadística, a la distribución de frecuencias de - - v.a. en hidrología.

Una F.D.P. que se ajuste a una distribución de frecuencias observadas, define las propiedades de localización (tendencia central, valores frontera), escala (dispersión, centralización), forma, asimetría, apuntamiento, asintotismo y otras propiedades; todas ellas valuadas con sus parámetros correspondientes. Desde el punto de vista matemático, a más parámetros en -- una función, mayor flexibilidad en el ajuste a la distribución empírica, y desde el punto de vista estadístico, la estimación de parámetros requiere, uno por cada momento, en algunos casos; lo que trae consigo rigidez en el cálculo, aunque en otros, los parámetros pueden estimarse directamente. Algunos de los parámetros estimados, serán menos confiables que otros, de donde surge la necesidad de optimizar; seguridad en los parámetros con-- tra flexibilidad en la función.

Es fácil convencerse, que en los ajustes de F.D.P. que toman solamente una distribución de frecuencias de una muestra, llevan en la mayoría de los casos a conclusiones equivocadas; y es aquí, donde juega un papel importante la experiencia del analista.

. Distribución normal de probabilidades.

La f.d.p., se expresa como:

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-(x-\bar{x})^2/2s^2} \quad ; \qquad \text{III.26.}$$

para $-\infty < x < \infty$. La correspondiente D.P.-A., es:

$$F(x) = P(X \leq x) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-\bar{x})^2/2s^2} dx;$$

III.27.

y X es la v.a., normalmente distribuída con media $\bar{x}$ y varianza s^2.
Si se hace:

$$Z = \frac{X - \bar{X}}{s}$$

donde Z es la variable normalizada correspon diente a X. El valor esperado para Z es cero y la varianza 1; luego entonces, la f.d.p. - normal tipificada queda:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad ; \qquad \text{III.28.}$$

y la correspondiente D.P.A., se expresa:

$$F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du \; ; \qquad \text{III.29.}$$

que en forma equivalente es:

$$F(z) = P(Z \leq z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-u^2/2} du \; ;$$

III.29'.

Observese en la fig. 3., que la integral anterior, representa el área bajo la curva normal a la derecha de cero.

Como anteriormente se expuso, una expresión para e^{-x} , de las series de Taylor, es de la forma:

$$e^{-x} = 1-x + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots - \frac{x^{n-1}}{(n-1)!} + \frac{x^n}{n!} \; ; \qquad III.30.$$

Si $x = \frac{u^2}{2}$, desarrollando y simplificando la solución de la integral de referencia líneas arriba, queda:

$$\int_0^z e^{-u^2/2} du = \sum_{j=o}^{z} \frac{u^{(2j+1)}}{(2j+1).2^j.J!} (-1)^j \; ; \qquad III.30'$$

y la expresión III.29'., resulta:

$$P(Z \leq z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left[\sum_{j=o}^{z} \frac{u^{(2j+1)}}{(2j+1).2^j.J!} (-1)^j \right] \; ;$$

III.31.

Algunas propiedades de la distribución normal son:

media $= \bar{x}$

varianza $= s^2$

desv. típica $= s$

coef. de sesgo $= \alpha_3 = \frac{m3}{s3} = 0$

coef. de curtois $= \alpha_4 = \frac{m4}{s4} = 3$

función generatriz de momentos:

$$M(t) = e^{-\bar{x}t} + (s^2t^2/2)$$

El valor de la sumatoria en III.31.; se re-- suelve para valores absolutos menores ó iguales que $\underline{E}$ previamente establecido.

La curva de densidad de probabilidad normal, f(x) y su curva de distribución de probabilidad acumulada F(x), con μ =o y tres valores de σ, σ = 0.5, 1.0 y 2.0 se representa en la fig. 3

La función de distribución normal standard, ocupa una posición inigualable entre las funcio nes de distribución en la teoría de la probabilidad, la matemática estadística y los procesos estocásticos.

Frecuentemente, se usa como patrón para comparar con otra (s) función (es) de distribución, pués algunas de estas convergen a la función -- normal en condiciones particulares.

La función normal, está definida y desarro-- llada prácticamente para valores reales de x. - El máximo es $x = \mu_s$, cada $f(u) = 1/\sigma\sqrt{2\pi} = 0.4/\sigma$. Los puntos de inflexión son para $\mu \pm \sigma$ con... $f(\mu \pm \sigma) = \sqrt{e}/\sigma\sqrt{2\pi}$

La x es asintótica y f(x) es simétrica cerca de $x = \mu$; el área bajo la curva es la unidad, - la media μ, la mediana $\underline{M}$, y la moda $\underline{m}$; son - - idénticas. El coeficiente de variación está definido solamente para $\mu \neq 0$, ó particularmente para $\mu > 0$.

La medida del error en la desviación σ_p, está definido como $\sigma_p = X_1 - \mu$ ó $\sigma_p = \mu - X_2$, con X_1 y X_2 los valores de x, simétricamente locali zados alrededor de μ, el área entre ellos de - la curva de densidad de probabilidad es 0.5; en este caso, con $\sigma_p = \pm 0.6765 = \pm 2\sigma/3$.

El teorema del límite central indica que la suma ó el promedio de las variables aleatorias, tiende a la distribución normal, como el número de variables se incremente.

La f.d.p. normal, es útil particularmente - en hidrología en algunos casos:

1.- Para variables hidrológicas aleatorias - con distribución simétrica empírica de - frecuencias.

2.- Como una f.d.p. en el análisis de erro-- res aleatorios.

3.- A manera de comparación con otra distri- bución.

4.- Cuando varios parámetros estadísticos -- pueden ser ajustados aproximadamente a - los de una distribución normal.

5.- Se usa en el método de Monte Carlo (GENE RACION DE DATOS) para simulación, donde primero se tendrían números aleatorios - normales, dependientes ó independientes, para después transformar a otro tipo de distribución.

La función normal, puede ser relativamente - adecuada a la distribución empírica de precipi- tación anual, de un número grande de registro.- Sin embargo, el propósito de ajuste, no es bue- no para lluvia mensual y escurrimiento anual, - así como para valores diarios; sin que lo ante- rior pueda siquiera, generalizarse.

Dos factores físicos justifican la aplicación de la función normal en hidrología.

Primero.- varias variables pueden ser cero para un intervalo grande de tiempo. El valor cero, tiene una probabilidad finita y la probabilidad acumulada para $x = 0$, puede interpretarse teóricamente, como la integral de la -- densidad de probabilidad para valores negativos, ($-\infty$ a cero) y la cola negativa puede - interpretarse como una parte imaginaria de la distribución normal.

Segundo.- valores negativos pueden ocurrir en algunas variables hidrológicas, tal es el caso de ríos largos, con pendiente superficial pequeña, de tal forma, que un escurrimiento - rápido ó un tributario, ocasionan la formación de remansos y entonces existe, aguas --- arriba, flujo negativo; con lo cual se justifica el uso de colas negativas de la función normal de probabilidad.

La función normal puede ajustarse a datos, -- analítica ó gráficamente; usando una escala de -- probabilidades se tendrá una línea recta en un -- gráfico en el cual el eje horizontal, es de probabilidades. Es evidente así mismo, que puntos a la izquierda de $F(x) = 0$ tienden a $-\infty$ y puntos a - la derecha de $F(x) = 1$, tienden a $+\infty$.

Usualmente, para estimar parámetros de funciones de distribución, se tienen: el método de mínimos cuadrados, método de los momentos, método de la máxima verosimilitud y el método gráfico. Los tres primeros, dan la media y variancia de la función normal, con aproximadamente la misma exacti-

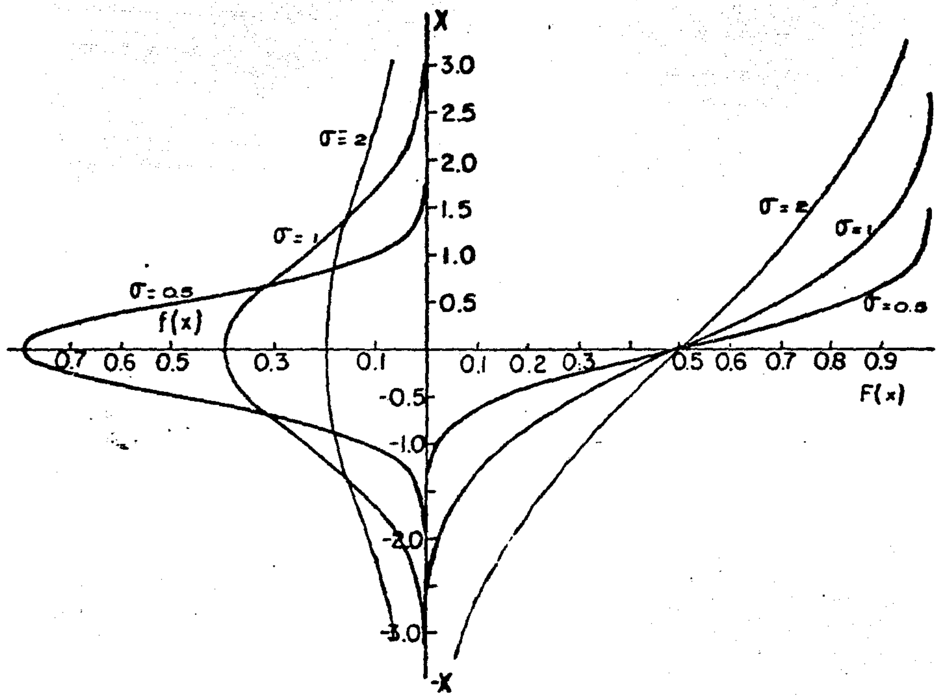


Fig. 3. Curva normal de densidad de probabilidad (lado izq. de la gráfica), y curva normal de distribucion de probabilidad (lado der. de la gráfica) para $\mu = 0$ y tres valores de σ, 0.5, 1.0 y 2.0

tud. La media μ es estimada por la media muestral $\bar{x}$, y la desviación standard σ es estimada por la desviación muestral, usando $\hat{s}$ para muestreos pequeños y s para muestreos grandes.

Adicionalmente, para determinar la bondad de ajuste, puede auxiliarse de Cs. (coeficiente de sesgo) y del coeficiente de curtosis.

. F.D.P. Logarítmico - normal.

El LnX, donde x está normalmente distribuida, se dice que tiene F.D.P. logarítmico-normal.

Haciendo:

$$f(x) = \varphi(LnX) \quad |d(LnX)| \qquad ; \text{III.32.}$$

y si: $$|d(LnX)| = \left|\frac{dx}{x}\right|, \text{ entonces:}$$

$$f(X) = \varphi\frac{(LnX)}{x} \qquad ; \text{III.32'.}$$

de donde se tiene la F.D.P. logarítmico-normal de la variable LnX de la siguiente forma:

$$\varphi(LnX) = \frac{1}{\sigma n\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{LnX - \mu n}{\sigma_n}\right)^2\right]; \qquad \text{III.33.}$$

y por consecuencia:

$$f(X) = \frac{1}{X\,\sigma_n\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{LnX - \mu n}{\sigma_n}\right)^2\right]; \qquad \text{III.33'.}$$

válida para $x \geq 0$, y donde μ_n y σ_n son la media y desviación standard de LnX.

La función Lognormal posee como límite inferior Cero y tiene un máximo único, los puntos de inflexión están dados por:

$$\mu_n - (3\ \sigma_n^2/2) \quad y \quad (\ \sigma_n\ \sqrt{4-\ \sigma_n^2/2}).$$

La curva es convexa hacia arriba entre los puntos de inflexión, y convexa hacia abajo en otros lados. La media, la mediana y la moda de esta función, son diferentes y frecuentemente se denota como:

$$L\ (\ \mu_n,\ \ \sigma_n).$$

La relación entre los parámetros μ , σ y μ_n, σ_n , permite obtener unos a partir de otros. Los momentos de X , alrededor de cero pueden obtenerse de:

$$\gamma_r = e^{(r\ \mu_n + r^2\ \sigma_n^2/2)} \quad ; \qquad III.34.$$

donde r es el orden del momento; así la media μ será con r=1 , y resulta:

$$\mu = e^{(\mu_n + \ \sigma_n^2/2)} \quad ; \qquad III.34'.$$

y de igual forma, la variancia de X es:

$$\sigma^2 = \mu^2(\ e^{\sigma_n^2} - 1) \quad ; \qquad III.34''.$$

por consecuencia, el coeficiente de variación resulta:

$$\eta = \frac{\sigma}{\mu} = (\ e^{\sigma_n^2} - 1)^{1/2} \quad ; \qquad III.35.$$

Frecuentemente es útil espresar μ_n y σ_n en términos de μ y σ ó μ y η ; ó μ y η . La relación entre sus respectivas expresiones dá como resultados:

$$\mu_n = \frac{1}{2} \operatorname{Ln} \left(\frac{\mu^2}{\eta^{2+1}}\right) = \frac{1}{2} \operatorname{Ln} \frac{\mu^4}{\mu^2 - \sigma^2} \; ; \qquad \text{III.36.}$$

y

$$\sigma_n^2 = \operatorname{Ln} (\eta^2 + 1) = \operatorname{Ln} \frac{\sigma^2 - \mu^2}{\mu^2} \; ; \qquad \text{III.37.}$$

Los coeficientes de sesgo y curtosis de X_i -- son respectivamente:

$$\gamma_1 = \eta^3 + 3\eta \qquad ; \qquad \text{III.38.}$$

y

$$\gamma_2 = \eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2 - 3 \; ; \qquad \text{III.39.}$$

así mismo, la moda y la mediana son respectivamente:

$$m = e^{(\mu_n - \sigma_n^2)} \qquad ; \qquad \text{III.40.}$$

$$M = e^{\mu_n} \qquad ; \qquad \text{III.41.}$$

Para determinar μ_n y σ_n, dos métodos son confiables, estimados μ y σ para los valores de la variable y usando III.36. y III.37., determinar μ_n y σ_n respectivamente; y estimar μ_n y σ_n directamente de los logarítmos de X_1 ó por el método de la -- máxima verosimilitud.

La fig. 4., muestra algunas curvas de la función lognormal para varios valores de μ_n y varianza σ_n^2 (y los correspondientes valores de μ ó σ^2).

Si la frontera inferior de la distribución lognormal no es cero, es necesario modificar la expresión III.33'. introduciendo el tercer parámetro y entonces:

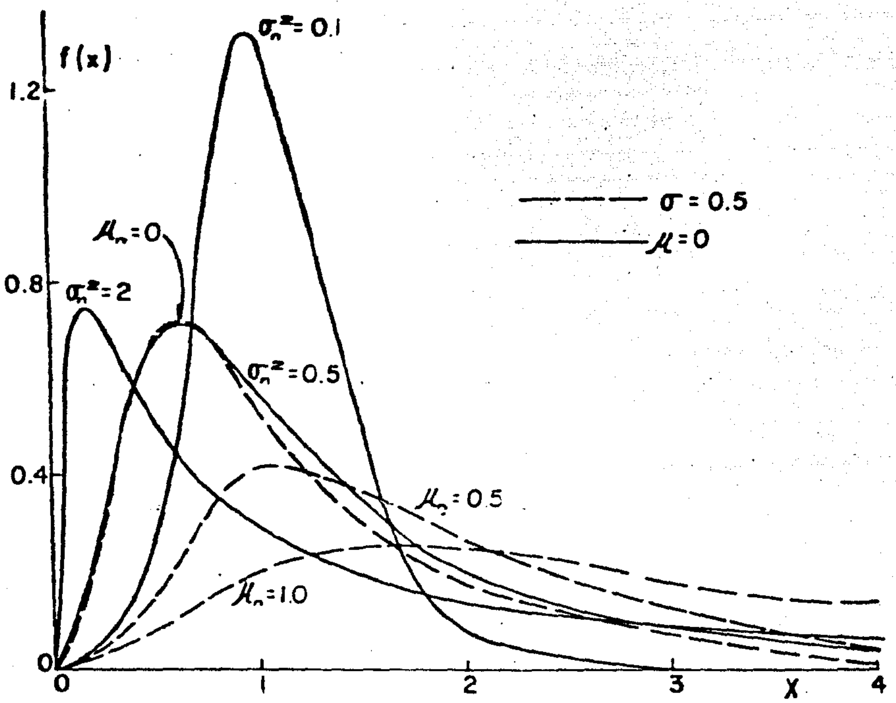


Fig. 4. Seis Curvas de probabilidad lognormal $\mu_n = 0, 0.5$ y 1.00, para $\sigma_n^2 = 0.5$ y $\sigma_n^2 = 0.1, 0.5$ y 2 para $\mu_n = 0$

$$f(x) = \frac{1}{(x-\beta)\ \sigma_n\sqrt{2\pi}} \exp\left\{ -\frac{\left[Ln(x-\beta) - \mu_n\right]^2}{2\ \sigma_n^{\ 2}} \right\};$$

III.42.

en la cual β representa la frontera inferior.

En el presente trabajo, se considera conocida la frontera inferior y la solución se aplica para III.33'., en igual forma que para la función normal de probabilidades, de tal suerte, que para -- considerar homogénea la notación, se tiene:

$$f(x) = \frac{1}{x\ Sn\ \sqrt{2\pi}}\ e^{-(Lnx-\bar{x}n)^2/2s_n^2} \quad ; \quad \text{III.42'.}$$

. F.D.P. Gamma con dos parámetros.

La f.d.p. es de la forma:

$$f(x) = \frac{1}{\beta^{\alpha}\ \Gamma(\alpha)}\ x^{\alpha-1}\ e^{-x/\beta} \quad ; \quad \text{III.43.}$$

para $0 \leq x < \infty$, pués $f(x)$ es igual a cero si $x < 0$.

Los parámetros α y β , son de forma y escala respectivamente y ambos mayores de cero.

La moda es: $m = \beta^{\alpha}(\alpha - 1)$; III.44.

para $\alpha \geq 1$, y $Ex = \mu = \alpha\beta$ es la media y var $x = \alpha\beta^2$, donde los coeficientes de sesgo y curtosis dados respectivamente por $\gamma_1 = 2/\sqrt{\alpha}$ y $\gamma_2 = 3\text{-}6\alpha$, son ambos independientes de β

En la fig. 5, tres curvas de III.43., se presentan para valores de $\alpha = 2$ y $\beta = 0.5$, 1.0 y 2.0.

Los parámetros α y β, son estimados de la muestra de datos. Para una distribución con sesgo alto, se recomienda usar el método de la máxima verosimilitud.

Anteriormente, para la F.D.P. Gamma se hizo que $\mu = \alpha\beta$ y $\sigma^2 = \alpha\beta^2$ de lo que resulta:

$$\beta = \sigma^2/\mu \qquad ; \qquad III.45.$$

y

$$\alpha = \mu^2/\sigma^2 \qquad ; \qquad III.46.$$

Por consecuencia lógica el coeficiente de variación C.v. $= \eta = 1/\alpha$; estimando μ y σ de $\bar{x}$ y $\hat{s}$. Esto es satisfactorio si el coeficiente de sesgo $\gamma_1 = 2/\sqrt{\alpha}$ es pequeño, ó α es relativamente grande, y si Cs y K son aproximadamente iguales a los valores de γ_1 y γ_2 respectivamente, enunciados líneas arriba. Debe observarse por lo anterior expuesto, que el valor de α puede usarse como un primer tanteo del ajuste, aunque al final, para valores "grandes" de (Cs $>$ 0.5), α, β y γ se estiman por el método de la máxima verosimilitud y la prueba de bondad del ajuste es la "chi cuadrado" que se enunciará en su oportunidad.

La D.P.A. es de la forma:

$$F(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} \int_0^{\alpha} x^{\alpha-1} e^{-x/\beta}\, dx \qquad ;III.47.$$

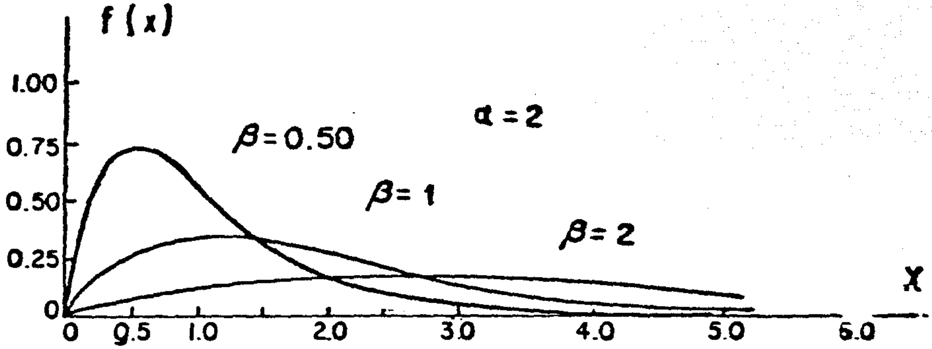


Fig. 5. Curva de densidad de probabilidad gamma de 2 parámetros, para α = 2 y tres valores de β; 0.5, 1, y 2.

y su respectiva solución, esto es, la de la inte-- gral de lado derecho de la expresión anterior será:

$$\int_0^{\infty} x^{\alpha-1} \; e^{-x/\beta} \, dx = \sum_{n=0}^{i} \frac{x^{(\alpha+n)}}{(\alpha+n)\beta^n n!} \; (-1)^n ; \qquad \text{III.48.}$$

y la sumatoria se resuelve para valores de:

$$E \geq 0.00001 .$$

Sin embargo, para resolver III.47. en su totalidad, falta la solución del término $\Gamma(\alpha)$ conocida como la función Gamma matemática, misma que - al término de este punto, se resolverá.

La F.D.P. Gamma, tiene la desventaja de no ser fácil su graficación en papel de probabilidad, de forma tal, que se tenga una línea recta y checar - un posible ajuste. Es conveniente hacer notar, que emplear la F.D.P. Gamma de tres parámetros no re-- presenta ventaja alguna sobre la de dos parámetros, pués el tercer parámetro es la frontera inferior y ésta función tiene utilidad para valores de $X \geq 0$.

. Función Gamma Matemática.

Función Gamma matemática ó función Gamma incompleta son términos equivalentes y la espresión que la define es:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \; e^{-x} \, dx \; ; \qquad \text{III.49.}$$

la cual puede demostrarse que converge para valo-- res de $\alpha > 0$. También se le llama función gamma de α ; y para valores de $\alpha > 0$ pero no ente-- ro, $\Gamma(\alpha)$ se calcule por serie e integración numérica, cuyo resultado es el siguiente:

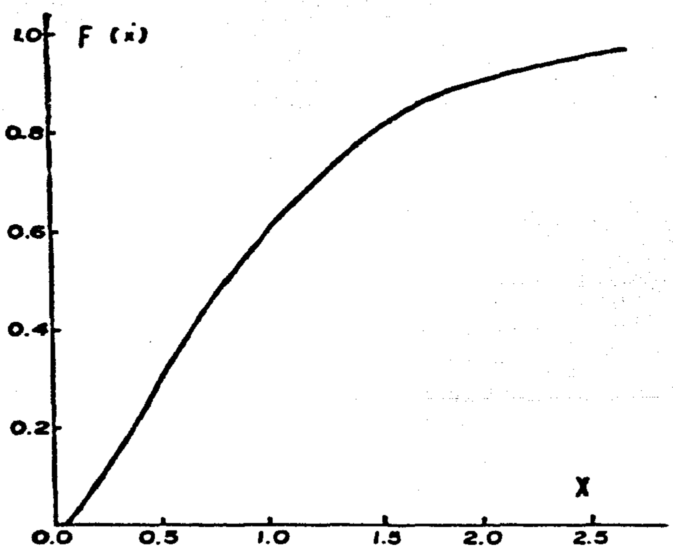


Fig. 6. Función de distribución gamma de 2 parámetros.

$$\int_0^{\infty} x^{\alpha-1} e^{-x}\, dx = \sum_{i=0}^{n} \frac{x^{(\alpha+n)}}{(\alpha+n)\; n!} (-1)^n \;;$$

III.50.

Nótese que si hacemos la expresión III.48., - $\beta = 1$ se tiene automáticamente III.50.

. D.P.A. De Valores Extremos.

Por la utilidad que representa para los propósitos hidrológicos, se expondrá aquí la F.D.P. de valores extremos máximos.

La función de distribución de probabilidad doble exponencial de extremos, esto es, valores máximos en su forma acumulada es:

$$F(x) = e^{-e^{-\alpha(x-\beta)}} \;; \qquad \text{III.51.}$$

en la cual F(x), es la distribución de probabilidad de los máximos valores de x de las submuestras, α y β son respectivamente parámetros de - escala y localización (valor central).

La media desviación standard y la mediana son:

$$\mu = \beta - \frac{0.5772}{\alpha} \;; \qquad \text{III.52.}$$

$$\sigma = \frac{\pi}{\alpha \sqrt{6}} \;; \qquad \text{III.53.}$$

$$M = \beta - \frac{0.3665}{\alpha} \;; \qquad \text{III.54.}$$

Para μ y σ conocidos, los parámetros α y β son:

$$\beta = \mu - 0.450\, \sigma \;; \qquad \text{III.55.}$$

$$\alpha = \frac{1.281}{\sigma} \;; \qquad \text{III.56.}$$

así mismo, los cuartes $F(X_1)$ y $F(X_2)$ son 0.25 y 0.75 respectivamente y α como una función de ellos, será:

$$\frac{1}{\alpha} = 0.6359\ (X_2 - X_1) \qquad \text{III.57.}$$

La relación inmediata anterior, puede em--- plearse como una primera estimación analítica de α

Los coeficientes de sesgo y curtosis son -- constantes, así $\gamma_1 = 1.29857$ y $\gamma_2 = 5.4$.

Para valores grandes de la frecuencia acumulada, $F(x) \geq 0.99$ la expresión III.51., se aproxima bien por la función $\phi(x) = e^{-y}$, la cual - es una función simple exponencial y donde . . .. $Y = \alpha\ (x - \beta)$ y puede considerarse de una cola - asintótica.

Una frecuente objeción hecha a la aplicación de la función doble exponencial a fenómenos con - una frontera inferior finita, es porque la fun--- ción es ilimitada. Cuando ésta función se aplica a valores positivos de la variable como en flujos pico, donde no puede ser cero, la función no es - estrictamente aplicable. Si una región es muy seca durante algunos años y los ríos no tienen escurrimiento, la probabilidad acumulada para la descarga cero es finita, lo cual puede justificar el cero de la función doble exponencial truncada.

Conocer la probabilidad para cada valor máximo en el análisis de flujo es de interés práctico en la hidrología.

III.3.- Prueba de ajuste y máxima verosimilitud.

Pueba de. ajuste.

Si se dispone de una muestra hidrológica de datos de tamaño "n", y se desea conocer sus propiedades estadísticas, la manera de lograrlo es aceptando a priori que dicha muestra tiene una cierta distribución de probabilidades conocida, y de ahí inferirlas, por lo que antes de proceder a utilizar y analizar dicha distribución, se requiere conocer qué tan cierto es que la distribución elegida se puede utilizar como representativa del conjunto de datos, ó muestra disponible.

Para éste propósito, se considera conveniente utilizar la prueba Chi cuadrado (χ^2), cuya aplicación adecuada, es a muestras grandes.

La función de distribución es de la forma:

$$p(\chi^2 \leq x) = \frac{1}{2^{\nu/2}\ \Gamma(\nu/2)} \int_0^x u^{(\nu/2)-1}\, e^{-u/2}\, du \qquad \text{III.58.}$$

y

$p(\chi^2 \leq x) = 0$ para $x < 0$, y donde ν es el número de grados de libertad.

Al observar III.58. y compararla con III.47, se nota que son exactamente iguales, salvo que $\alpha = \nu/2$ y $\beta = 2$. Por lo tanto, se dice que la D.P.A. (χ^2) es un caso especial de la D.P.A Gamma de 2 parámetros.

Si se consideran O_i y e_i respectivamente para los q intervalos, las frecuencias observadas del fenómeno y las esperadas teóricamente de acuerdo con la ley de distribución de probabilidades escogida como representativa del fenómeno; al aplicar la prueba X^2 , resulta:

$$X^2 = \sum_{i=1}^{q} \left[\frac{(O_i - e_i)^2}{e_i} \right] ; \qquad \text{III.59.}$$

donde:

q = número de intervalos.
m = número de parámetros que se deben estimar para determinar las frecuencias esperadas.

En la expresión III.58. los grados de libertad se expresan por: ν = q-m-1.

Al resolver III.58. para ν y nivel de significancia previamente establecido, se tiene un valor al que se denominará: "valor tabular". Y al resolver III.59. para las condiciones anteriores, se tendrá un valor al que denominaremos "valor calculado".

Si el valor calculado es mayor que el tabular, entonces la distribución elegida como de las frecuencias observadas de los datos analizados, no es la correcta y si sucede lo contrario, la distribución propuesta es aceptable, con un nivel de confianza.

Al utilizar esta prueba, como recomendación práctica, debe tenerse cuidado de que en cada intervalo de clase, se tengan por lo menos cinco datos observados.

* Debe observarse que la corrección de Yates puede emplearse cuando existe duda en los resultados; sobre todo en muestras pequeñas comparando X2 corregido y X2 no corregido.

Máxima verosimilitud

Sea

$$f(x;\alpha ,\beta =...)\_$$

una función densidad de probabilidad de x, con α β... , los parámetros a estimarse. El producto,

$$L = \prod_{i=1}^{N} f\ (Xi\ ;\ \alpha,\beta ...) \qquad III.60$$

Se llama función de verosimilitud de una muestra- de N valores para una población de variables contínuas. X. En el caso de variables discretas con- probabilidad acumulada, Pi (X;α,β...), la función de verosimilitud es

$$L = \prod_{i=1}^{N} Pi\ (Xi\ ;\alpha,\beta\ ,....) \qquad III.61$$

En la cual N es el tamaño de la muestra y α ,β, ... se sustituyen por los estimados a, b, ... La estimación deα,β, γ, consiste en determinar a, b, c, ... para la muestra de datos para lo cuál un camino es - obtener el máximo valor posible de L.

Como ln L alcanza el valor máximo al darle los valores a, b, c, ... la ecuación de verosimilitud es:

$$\ln L = \ln \prod_{i=1}^{n} f\ (Xi;\ a,\ b,\ ...\ ) = \sum_{i=1}^{n} \ln - f\ (Xi;\ a,\ b,\ ...\ ) \qquad III.62$$

Despues de tomar derivadas parciales en a, b, ...,-
é igualar a cero los resultados, las ecuaciones de
máxima verosimilitud son:

$$\frac{\partial \ln L}{\partial a} = 0 \; ; \; \frac{\partial \ln L}{\partial b} = 0 \; ; \ldots \qquad \text{III.63.}$$

Estas ecuaciones, de número igual al número de parámetros, pueden resolverse en forma simultánea y calcular a, b, c ..., para una muestra dada, donde los parámetros citados son estimadores eficientes de α, β...; si los estimadoreseficientes existen.

* Cuándo el análista considere necesario puede emplear el estadístico de SMIRNOV-KOLMOGOROV para la bondad del ajuste.

III.4.- Algoritmos de cálculo.

a) A efecto de salvar el obstáculo que representa tener valores menores de 1, en la -- función lognormal, después de ordenarlos - de menor a mayor, se incrementan los valores en el valor menor más una variable - - (RINC), a criterio del análista.

b) Todas las integrales de las D.P.A. se re-- suelven por series de Taylor y usando el - término general de cada serie, donde debe notarse que se avanza de 0 a n ; aunque pudiera hacerse de 1 a n, usando un artificio en el término general.

c) Para evitar cálculo engorroso, el facto--- rial se resolvió en una función (IFAC).

IV.- ESTRUCTURACION, DESARROLLO Y OPERACION DEL MODELO

IV.1.- Estructura básica y diagrama de bloques.

La estructura fundamental del modelo, descansa en tres puntos básicos:

a) Los datos, son el resultado de un análisis de muestreo, mediante el cual se definió tipo y magnitud de la muestra por analizar.

b) A priori se asigna una distrubución de probabilidad a los datos muestrales, y se calculan sus correspondientes parámetros estadísticos.

c) Se efectúan las pruebas de ajuste (X^2) y máxima verosimilitud.

Observe como el trabajo se resuelve para las "n" distribuciones implementadas al programa. Para mayor claridad, puede verse en la fig. 2, la estructura básica y en la fig. 7 el correspondiente al diagrama de bloques.

IV.2.- Estructura general y funcionamiento lógico

A efecto de darle utilidad practica a este trabajo, se ha procurado en la medida de lo posible, la mejor utilización de los cálculos y no exceder limites de tiempo de proceso de las corridas del programa. En primera instancia, se encuentran las lecturas de datos, el arreglo de menor a mayor para los mismos, el cálculo de los logaritmos naturales de ellos y el cálculo de parámetros estadísticos; consecuentemente, la asignación de distribución de probabilidades y al final de este primer proceso, la prueba de ajuste y resultados.

Lo anterioe, se consigue integrando en el bloque del programa principal. Los cálculos particulares y en subrutinas los comunes.

Así, el modelo está constituído por el programa principal TEST, subrutinas ELEM, GAMMA, TESTI y VEROS.

Lo anterior, se aclara en la figura 8, que muestra un listado del programa en lenguaje FORTRAN IV.

Una ventaja de este modelo, es que con relativa facilidad, podran incluirse en el programa, las distribuciones de probabilidad del usuario considere conveniente para enfocar su problema particular; aunque esto no implica que se modifique su funcionamiento lógico.

IV.3.- Operación

IV.3.1.- Datos de entrada

VARIABLE	FORMATO	SIGNIFICADO
NM	15	Número de muestras.
NPA	15	Número de parámetros
IHN	15	Año inicial.
NIN	15	Ensayos dif. intervalos
IN	15	Número de intérvalos.
ND	15	Número de datos.
RINC	F6.I	Incremento al mínimo valor muestral.
RNS	F6.I	Constante de Conversión.
DAT	F8.2	Datos muestrales.

IV.3.2.- R E S U L T A D O S

Estos son los valores de los parámetros estadísticos y los de la prueba de ajuste. La utilidad practica radica en la buena interpretación que de ellos haga el analista como ejemplo vease la figura 9.

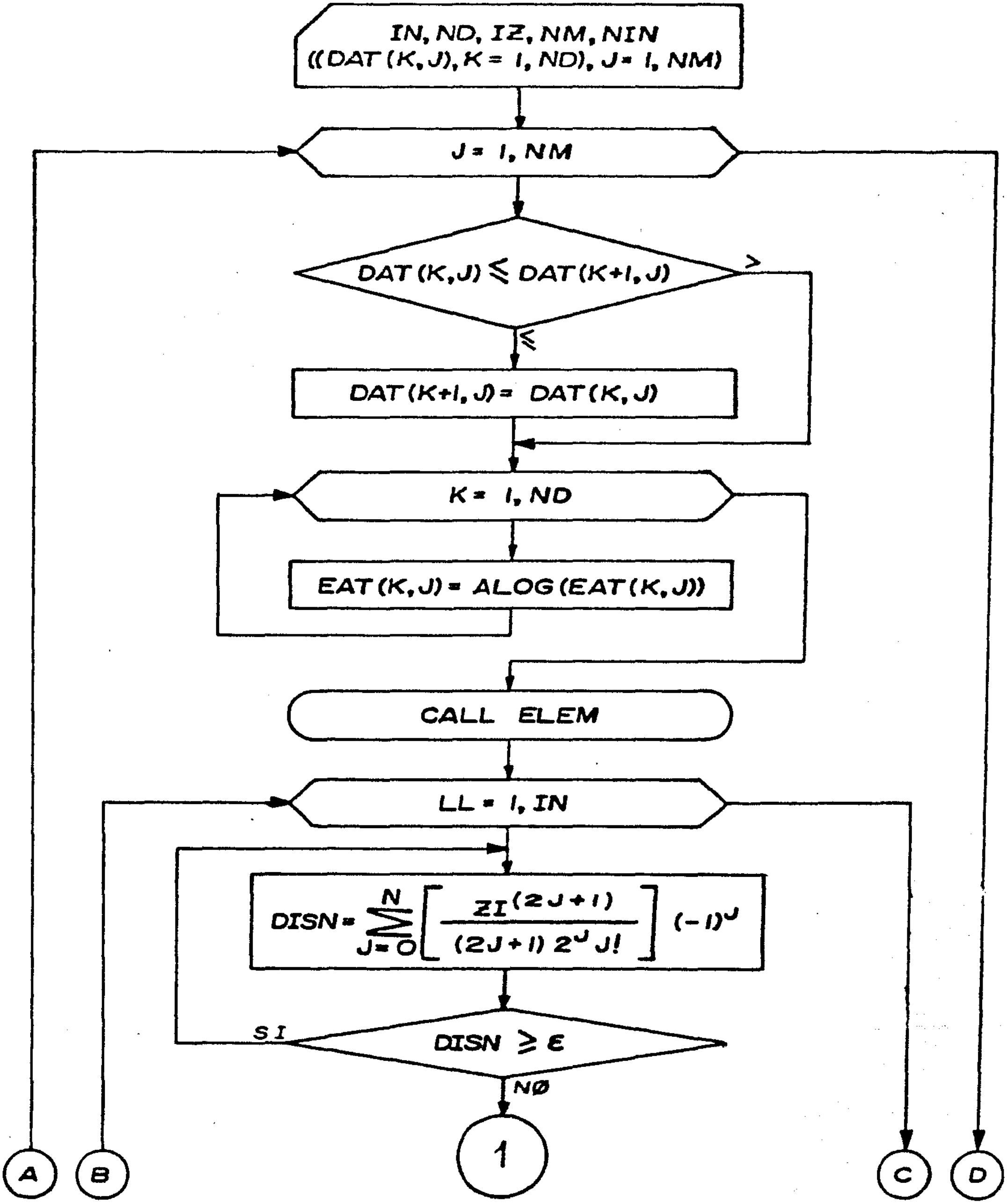
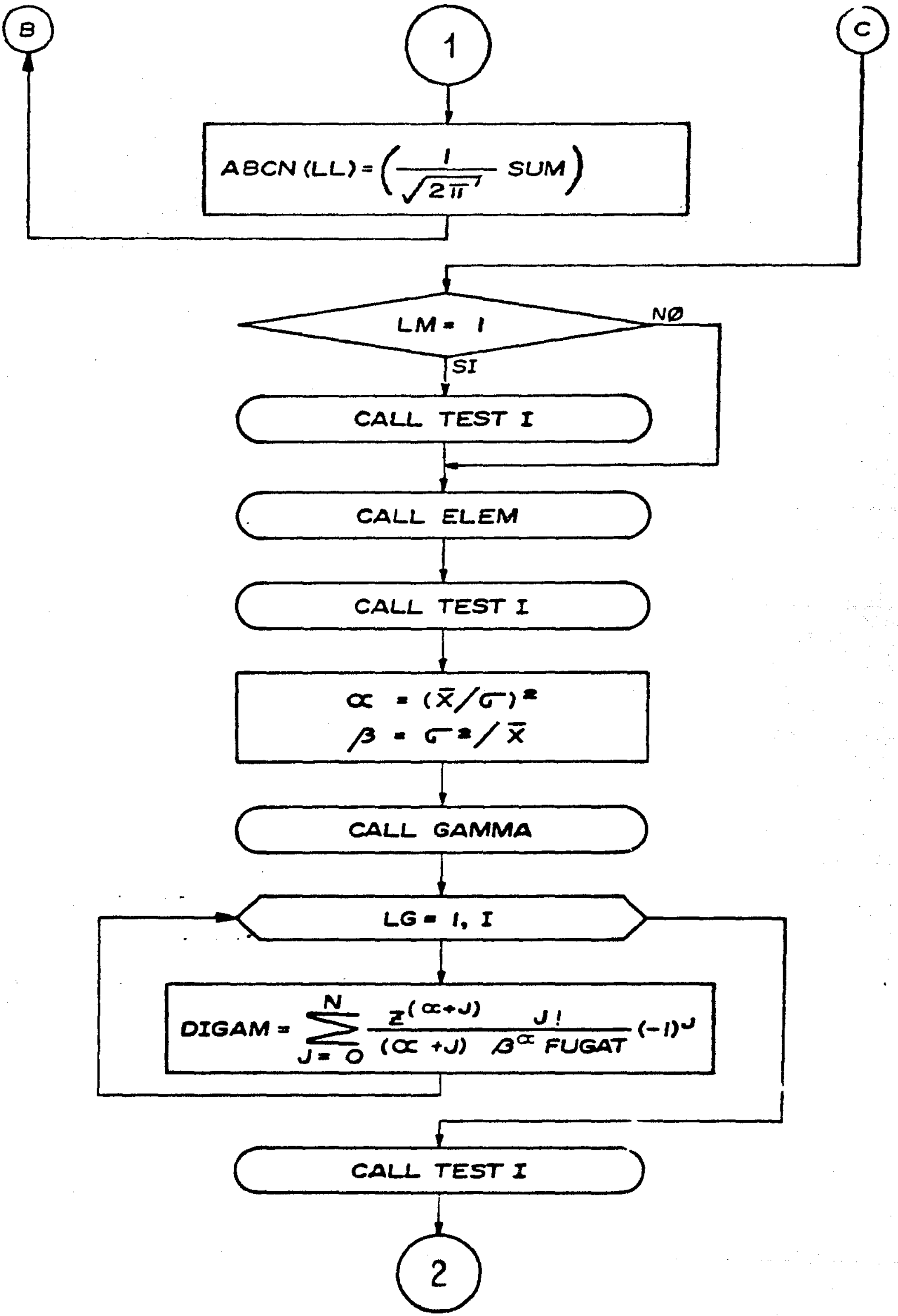
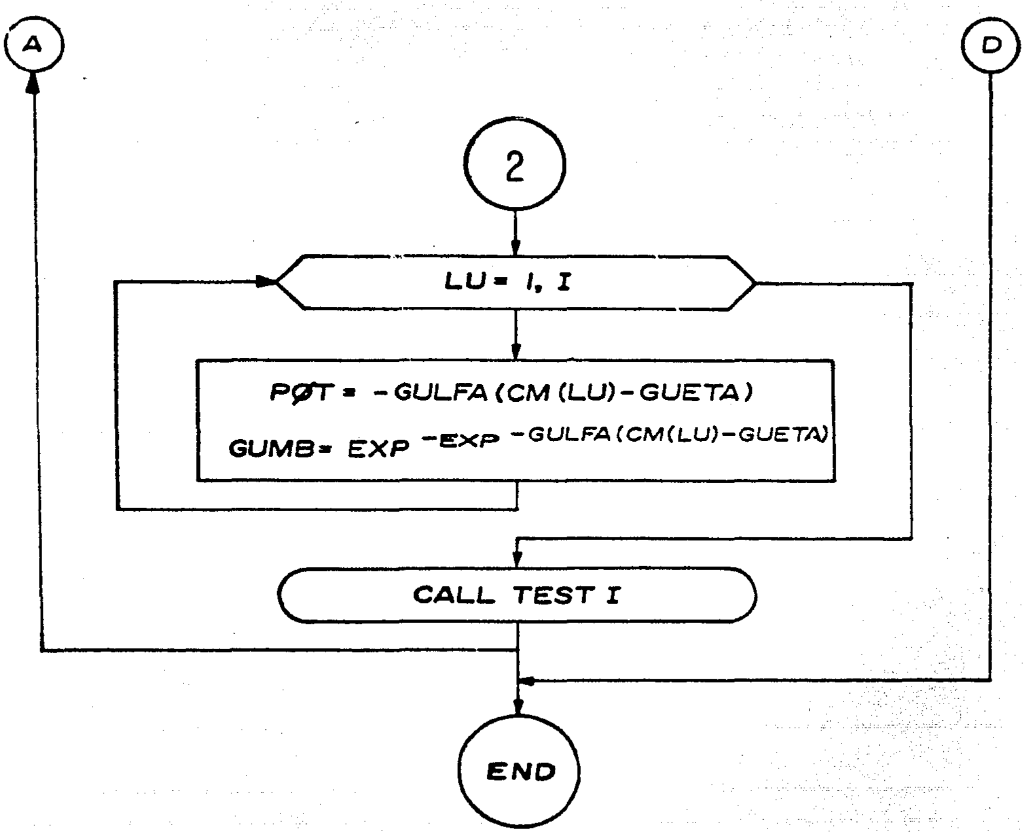


Fig. 7 DIAGRAMA DE BLOQUES DEL PROGRAMA TEST


B
1
C
ABCN (LL) = (1/√2π SUM)
LM = 1
NØ
SI
CALL TEST I
CALL ELEM
CALL TEST I
α = (X̄/σ)²
β = σ²/X̄
CALL GAMMA
LG = 1, I
DIGAM = Σ_{J=0}^{N} Z^(α+J)/(α+J) · J!/(β^α FUGAT) (−1)^J
CALL TEST I
2



A
D
2
LU = 1, I
PØT = -GULFA (CM (LU) - GUETA)
GUMB = EXP -EXP -GULFA (CM(LU)-GUETA)
CALL TEST I
END

IV.3.3.- Limitaciones.

. La solución se plantea para valores positivos de las variables.

. La eficiencia del modelo, puede ser en un momento dado, función de causas externas; como del análisis de muestreo, correcto enfoque del problema que se enfrenta, restricciones particulares del centro de cómputo donde se corra, etc..

. Que en el número de distribuciones analizadas no se encuentre la más adecuada a la distribución empírica.

. La estimación incorrecta de los parámetros estadísticos para un problema en particular.

V. EJEPLO DE APLICACION.

V.I.- I N T R O D U C C I O N.

En este ejemplo los resultados se muestran en la siguiente hoja y el fin único es verificación del funcionamiento del modelo, - dejando para sus posibles aplicaciones prácticas las deducciones que de ellos se originen; al comparar los parámetros empíricos de la muestra, los estadísticos muestrales, los resultados de las pruebas de ajuste y máxima verosimilitud con los parámetros esperados, para un caso particular.

V.2.- E J E M P L O.

Se tomó una muestra de 39 años (1902-1906) de gastos máximos -- anuales y cuyos resultados pueden correlacionarse con los que - se encuentran en el libro uno (1) de la bibliografía al final de este trabajo.

Así mismo, si se procede a la solución gráfica del problema utilizando papel de probabilidad, debe notarse bastante similitud con los resultados aquí presentados.

```
      PROGRAM  TEST  (INPUT, OUTPUT,  DEBUG = OUTPUT)



      PROGRAMA TEST

      DEFINICION  DE  VARIABLES

      DAT   = DATOS   MUESTRALES
      RINC  = INCREMENTO IGUAL AL MINIMO VALOR MUESTRAL
      IN    = INTERVALOS DE CLASE.
      ND    = NUMERO DE DATOS POR MUESTRA
      NM    = NUMERO DE MUESTRAS
      RNS   = CONSTANTE DE CONVERSION.
      NIN   = ENSAYOS   DIF NUM INTER.
      MEDA  = MEDIA  ARITMETICA
      VMO   = MODA
      VM    = MEDIANA
      VAR   = VARIANCIA
      DEST  = DESVIACION
      CV    = COEFICIENTE  DE  VARIACION
      COVA  = COEF  DE VARIACION
      COAS  = COEF  DE ASIMETRIA
      CODE  = COEF.DE SESGO.
      VMK   = COEF  DE CURTOSIS
      SUBROUTINE ELEM    :  ELEMENTOS EST. MUESTRALES.
      SUBROUTINE TESTI   :  PRUEBA  DE AJUSTE CON JI CUADRADO.
      SUBROUTINE GAMMA   :  SOLUCION  DE FUNCION GAMMA MAT.
      SUBROUTINE TESTMC  :  PRUEBA DE AJUSTE CON MIN. CUADRADOS.
      SUBROUTINE PARES   :  PARAM. EST.DIST.NORMAL ASIGNADA.

      DIMENSION   AMIN( 60 ),  ZI(60 ) ,  ABON( 60 ). ADIN( 60 ), FRES(
     160 ), GI (60)  , ABOG(60 ), AGUDI( 60 ), FREG(60 ), DAT(600,10),
     2DAT(600, 10), CM(60 ), FREGU( 60 ), FOBS( 60 ),FREL(60),FRELA(60),
     3XLN (60) ,FRESR(60),FRESRA(60),ADGA(60),FRESG(60),FRESGA(60),ZII(6
     40),GII(60) ,SUM(60)
      COMMON AMI(60)
      READ 5 , NM, NFA, IAN ,NIN
      PRINT  30
      DO 440   L = 1,  NM
      DO 410  LN = 1, NIN
      READ  10, IN, ND, RINC, RNS
      READ  20, (DAT(K,L) , K=1, ND)
      DO 25 IM =1, ND
      IAN = IAN+1
      DAT(IM,L) = DAT(IM,L) / RNS
      PRINT 35, IAN,DAT (IM,L)
   25 CONTINUE
      K=1
```

```
24 IF( DAT(K,L ).LE. DAT(K+1,L))29, 27
27 DATO    DAT  (K+1,L)
   DAT (K+1, L) = DAT(K,L)
   DAT (K, L) = DATO
   K = 
   GO  TO  24
29 K = K+1
   IF( K.LE. ND-1) GO TO 24
   IF( DAT(1,L) .LE. 1.0  )  DAIN = DAT (1,L) + FINC
   PRINT 9
   DO 31  K = 1, ND
   EAT(K,L) = ALOG (DAT(K,L))
   DAT(K, L)  = DAT(K,L) + DAIN
   PRINT 15  , DAT(K,L ), EAT( K,L)
31 CONTINUE
   RAN = DAT( ND,L) - DAT(1,L)
   PRINT 10
   PRINT 11  ,    DAT(ND,L) , DAT(1 ,L) , RAN
   RANL  = EAT(ND,L)  -  EAT(1,L)
   PRINT 11 , EAT(ND,L), EAT(1,L), RANL
   PRINT 12
   LI=1
   CALL ELFM ( IN, DAT, ND, CM, DEST, VAR, LI, RMEDA, FOBS,AMIN,L)
   PRINT 12
   FREL(1) =  FOBS(1) /ND
   FRELA (1) =  FREL (1)
   DO 136 MI = 2, IN
   FREL (MI)  = FOBS(MI) / ND
   FRELA (MI) =  FREL(MI) + FRELA(MI-1)
136 CONTINUE
   DO 14  MI =1, IN
   PRINT 14,MI,AMI (MI),AMI  (MI+1),CM(MI),FOBS(MI),FREL(MI),FRELA(MI)
14  CONTINUE
   LM=1
   GO TO 145
15  LM = LM+1
145 CONTINUE

   DO 158 LL=1, INN
   ZI (LL) = (AMIN (LL)  - RMEDA ) / DEST
   IF(LM.EQ.2) GO TO 156

   GO TO 15

15  SUM = 0.0
   ZII(LL)=ABS(ZI(LL))


   IF(KJ.EQ.0) KJ =1
165 DISN=(((ZII(LL)**(2*J+1))/((2*J+1)*2**J*IFAC(KJ))))*(-1)**J
   SUM = SUM+ DISN
```

```
      J = J+1
      KJ= J
      IF( ABS(DISN).GE...0    ) GO TO 165
      ABCN (LL) =   ( SUM /(2.50663*XLN(LL)))
  166 CONTINUE
      DO 180 LL=1, IN
      SIGNO =   ZI(LL+1)/ZI(LL)
      IF( SIGNO  .GT...) GO TO 175
      ADIN(LL)   =       ABCN  (LL+1)+ABCN(LL)
      GO TO 176
  175 ADIN(LL) =          ( ABCN(LL+1)-ABCN(LL))
  176 FRES(LL) =   ADIN(LL) * ND
      FRET=FRES(LL)+FRET
  180 CONTINUE

      PRINT 36
      IF(LM.EQ.1)PRINT 170
      IF(LM.EQ.2 )PRINT 173
      FRESR(1 )   = FRES (1) / FRET
      FRESRA(1)   = FRESR(1)
      DO 189 MM = 2, IN
      FRESR (MM) =  FRES(MM) / FRET
      FRESRA(MM) =  FRESR(MM) + FRESRA(MM-1)
  189 CONTINUE
      DO 192  N = 1, INN
      PRINT  15, N, ZI(N), ABCN(N), ADIN(N), FRES(N),FRESR(N),FRESRA(N)
  192 CONTINUE
      IF(LM.EQ.1) CALL PARES (IN, FRES,OM,IFRET)
      NU =  IN-NPA-1
      IF(LM.EQ. 1) CALL  TESTI(IN,FOBS,FRES, LM, NU)
      IF(LM.EQ.2) GO TO 196
      CALL ELEM  (IN, DAT, ND, CM, DEST, VAR,LM,RMEDA,FOBS,AMIN,L)
      IF(LM.EQ.1) GO TO 130
  196 XMELN  =2.7182**(RMEDA +DEST**2 /2)
      XDEL  =(XMELN **2) * (2.7182**  (DEST**2) +1)
      XDELN =  SQRT(XDEL)
      COVLN = XDELN / XMELN
      SESLN =  (COVLN **3)  + 3*COVLN
      CURLN =(COVLN**8)+6*(COVLN**6)+15*(COVLN**4)+16*(COVLN**2)-3
      PRINT 198, XMELN, XDELN, COVLN, SESLN, CURLN
      CALL TESTI( IN, FOBS, FRES, LM, NU)
      CALL ELEM ( IN, DAT, ND, CM, DEST, VAR, LM, RMEDA, FOBS,AMIN,L)
      ALFA  = (RMEDA/ DEST )**2
      BETA = DEST ** 2 / RMEDA
      CALL    GAMMA ( ALFA  , FUGAT)
      RALFA = SQRT (ALFA)
      GMED  = ALFA * BETA
      GDES  = BETA * RALFA
      SESGA  =   2.0*SQRT(ALFA)
      CURGA  =   3.0   - 6 * ALFA
      FRCD  = 1.0 /((BETA**ALFA)*FUGAT)
```

```
      DO 275  LG= 1, IN
      GI(LG)=(AMIN(LG)- GMED) /GDES
      SUG = 0.0
      GII(LG)=ABS(GI(LG))
      J    = 0

      IF(KJ.EQ.0) KJ =1
  205 DIGA=(GII(LG)**(ALFA+J)/((ALFA+J)*BETA**J*IFAC(KJ)))*(-1)**J
      SUG = SUG + DIGA
      J = J+1

      IF( ABS(DIGA). GE..00     ) GO TO 205
      ABCG(LG) = PROD * DIGA
      IF (LG.EQ.1) GO TO 245
      SIGA = ( GI(LG) /GI(LG-1))
      IF (SIGA.GT..0)GO TO 225
      ADGA(LG) =  ABCG(LG) + ABCG(LG-1)
      GO TO 255
  225 ADGA(LG) =     (ABCG(LG)-ABCG(LG-1))
      GO TO 255
  245 ADGA(1) = ABCG(1)
  255 FREG(LG) =ADGA(LG) *ND
  275 CONTINUE
      LM = LM+1
      PRINT 36
      FRESG(1) = FREG(1) /ND
      FRESGA(1) = FRESG(1)
      DO  285 IG = 2, IN
      FRESG(IG) = FREG (IG) /ND
      FRESGA(IG) = FRESG(IG) + FRESGA(IG-1)
  285 CONTINUE
      PRINT 290
      DO 295 NG=1, IN
      PRINT 15, NG, GI(NG),ABCG(NG),ADGA(NG),FREG(NG),FRESG(NG),
     1FRESGA(NG)
  295 CONTINUE
      PRINT 297
      PRINT 299, GMED, GDES, SESGA, CURGA
      CALL  TESTI ( IN,FOBS,FREG,LM,NU )
      GULFA = 1.281 / DEST
      GUETA  = RMEDA - 0.45 * DEST
      GUMENA  =  GUETA - (0.5555 / GULFA )
      DO  330  LL = 1, IN
      POT  = (-1) * GULFA * (GM (LL) - GUETA)
      GUM(LL) =  2.7182 ** (-2.7182 ** POT)
  330 CONTINUE
      PRINT 305, GULFA, GUETA, GUMENA
      CALL  TESTMO ( IN, FOBS, GUM, LM )
  400 CONTINUE
```

VI. CONCLUSIONES.

Básicamente, la selección de una distribución conveniente para representar una variable aleatoria cuando la única fuente de - información son los datos empíricos observados; es un proceso en tres etapas:

1.- La primera etapa, es una apreciación inicial cualitativa de las distribuciones posibles para eliminar las que no reproducen características globales importantes de los datos. En este caso, la exactitud del ajuste es el criterio que rige la elección.

2.- La segunda etapa, consiste en emplear un procedimiento cualitativo de "estimación del modelo". Al respecto, - es bastante razonable el método de máxima verosimilitud.

3.- La etapa de elección final es aquella en la cual se deben manifestar capacidad, experiencia relativa de facilidad de cálculo, exactitud de la representación y falla de un modelo al reproducir aspectos aparentes de - los datos.

Debe observarse un proceso iterativo que permita una revisión - también iterativa.

En términos generales, el presente modelo es una herramienta - útil para resolver problemas que onvolucren de una u otra forma variables aleatorias, iniciando la solución con un concienzudo análisis de muestreo; donde los estadísticos muestrales - pueden ser medidas de centralización, dispersión o selección - de valores máximos o por cierto intervalo de tiempo de ocurrencia. A continuación y en base a lo anterior, asignar las distribuciones implementadas en el modelo a las muestras obteni--

das de observaciones.

Los parámetros para cada distribución pueden obtenerse por el método de máxima verosimilitud; así por ejemplo; para la distribución normal los mejores estimadores de centralización y dispersión son la media y la desviación comose señala en las ecuaciones III.8 y III.9 respectívamente, obtenidos de la -- función de probabilidad respectiva.

Para plantear la regla de descisión deben establecerse; hipótesis nula, alternativa y niveles de confianza, pero sobre todo; el usuario del presente modelo debe justificar su uso y - tener la certeza de haber enfocado correctamente el problema al que se enfrenta.

PROGRAMA TEST

EJEMPLO DE APLICACION :

SELECCIONAR UN MODELO PARA INFERIR INFORMACION RESPECTO A LAS OBSERVACIONES DE LOS CAUDALES MAXIMOS ANUALES EN UN RIO, CUYOS REGISTROS DE 1902 A 1961 SON :

ANIO	DATO MUESTRAL (FT3/SEG EN MILES)
1902	45.000
1903	102.000
1904	118.000
1905	81.000
1906	128.000
1907	165.000
1908	16.300
1909	140.000
1910	31.000
1911	75.400
1912	16.400
1913	16.300
1914	122.000
1915	81.400
1916	42.400
1917	80.400
1918	28.200
1919	65.900
1920	23.400
1921	62.300
1922	36.400
1923	22.400
1924	42.400
1925	64.300
1926	55.700
1927	94.000
1928	195.000
1929	14.000
1930	80.100
1931	11.600
1932	22.600
1933	8.360
1934	20.300
1935	58.600
1936	35.400
1937	19.200
1938	195.400
1939	3.380
1940	152.000
1941	94.200

1942	12[illegible].000
1943	108.000
1944	24.900
1945	60.010
1946	54.400
1947	45.500
1948	36.700
1949	16.800
1950	46.400
1951	32.100
1952	13.000
1953	113.000
1954	54.500
1955	59.200
1956	203.000
1957	83.100
1958	102.000
1959	34.500
1960	135.000

DATOS MUESTRALES DE MENOR A MAYOR Y LNDATO

DATOS MUESTRALES	LN DATO
8.730	2.169
8.860	2.182
11.600	2.451
13.000	2.565
14.000	2.639
16.300	2.791
16.400	2.797
16.800	2.821
16.800	2.821
19.200	2.955
20.300	3.011
22.400	3.109
22.600	3.118
23.400	3.153
24.900	3.215
28.200	3.339
31.000	3.434
34.500	3.541
36.400	3.595
36.700	3.603
40.000	3.689
42.400	3.747
42.400	3.747
45.600	3.820
46.400	3.837
54.400	3.996
54.800	4.004
55.700	4.020
58.600	4.071
59.200	4.081
60.800	4.108
62.300	4.132
64.300	4.164
65.900	4.188
75.400	4.323
80.100	4.383
80.400	4.387
81.000	4.394
81.400	4.399
83.100	4.420
84.200	4.433
85.400	4.447
92.100	4.523
94.000	4.543
102.000	4.625
102.000	4.625
108.000	4.682
110.000	4.700
113.000	4.727
118.000	4.771
122.000	4.804
128.000	4.852
135.000	4.905

14[illegible].0[illegible]	4.942
152.0[illegible]	5.324
165.[illegible]	5.1[illegible]6
185.0[illegible]	5.223
185.[illegible]	5.223
2[illegible]3.0[illegible]	5.313

XMAX = 203.0   XMIN =   8.1

RANGO = 194.9

XMAX =   5.3   XMIN =   2.1

RANGO =   3.2

PARAMETROS ESTADISTICOS DE LA DISTRIBUCION MUESTRAL EMPIRICA

MEDI X= 69.1

DESTX = 47.9

MEDIAN= 59.2

MODAX = 39.5

COEF VAR =   .7

COEF ASIME=   .6

SESGO   =   .6

CURTOSIS   =   2.8

FRECUENCIAS OBSERVADAS

	INT CLASE	MAR CLASE	FREC	FRE REL	FREC REL AC
1	8.18 - 24.32	16.25	14.0	.2373	.2373
2	24.32 - 40.57	32.45	7.0	.1186	.3559
3	40.57 - 56.81	48.69	7.0	.1186	.4746
4	56.81 - 73.05	64.93	6.0	.1017	.5763
5	73.05 - 89.30	81.18	8.0	.1356	.7119
6	89.30 - 105.54	97.42	4.0	.0678	.7797
7	105.54 - 121.78	113.66	4.0	.0678	.8475
8	121.78 - 138.03	129.91	3.0	.0508	.8983
9	138.03 - 154.27	146.15	2.0	.0339	.9322
10	154.27 - 170.51	162.39	1.0	.0169	.9492
11	170.51 - 186.76	178.64	2.0	.0339	.9831
12	186.76 - 203.00	194.88	1.0	.0169	1.0000

PRUEBA DE AJUSTE

DISTRIBUCION : NORMAL

FRECUENCIAS ESPERADAS

X	Z	F(X)1	P(X)R	FESP	FREC	FRECA
1	-1.28	[illegible]	.0611	4.8	.1903	.1913
2	-.91	.3103	.1023	6.1	.1139	.2142
3	-.57	.2171	.1242	7.3	.1332	.3424
4	-.23	.0925	.1343	7.9	.1495	.4921
5	.1[illegible]	.0415	.1297	7.7	.1444	.6364
6	.44	.1712	.1120	6.5	.1246	.7610
7	.78	.2871	.0955	5.1	.0951	.8561
8	1.12	.3686	.0598	3.5	.0666	.9227
9	1.46	.4279	.0358	2.1	.0339	.9623
10	1.80	.4641	.0197	1.2	.0219	.9845
11	2.14	.4838	.0087	.5	.0097	.9941
12	2.48	.4926	.0053	.3	.0059	1.0000
13	2.82	.4979	[illegible]	[illegible]	[illegible]	[illegible]

PARAM EST. ESPERADOS DE LA DIST.NORMAL ASIGNADA

VEMED = 77.130

VESTA = 1592.379

CVESP = 20.639

VESGO = .000

VESK = .000

RESULTADOS DE LA PRUEBA X2(JI CUADRADO)

DISTRIBUCION NORMAL

NIV CONF = 5.00

JI X2 TABULADO = 21.0300

INT	FRES OBS	FRES ESP	ERROR
1	14.000	4.786	17.738
2	7.000	6.036	.154
3	7.000	7.327	.015
4	6.000	7.926	.468
5	8.000	7.654	.016
6	4.000	6.606	1.028
7	4.000	5.042	.215
8	3.000	3.538	.079
9	2.000	2.113	.006
10	1.000	1.161	.022
11	2.000	.512	4.321
12	1.000	.311	1.528

JI X2 CALCULADO = 25.5891

HO : CHIC = CHIT

H1 : CHIC MENORQUE CHIT

SI H1: SE ACEPTA EL AJUSTE COMO BUENO AL : 5.00DE CONFIANZA :

PRUEBA DE AJUSTE

DISTRIBUCION : LN NORMAL

FRECUENCIAS ESPERADAS

Y	Z	P(Y)ALN	F(X)RLN	FESPLN	FRERLN	FRERALN
1	-2.38	5.5983	[illegible]	492.6	.8337	.8337
2	-.72	.5395	.3164	16.7	.0316	.8653
3	-.39	.2721	.1973	11.6	.0197	.8851
4	-.06	.1246	.116	6.3	.0106	.8956
5	.27	.0812	.0426	2.5	.0043	.8999
6	.60	.1273	[illegible]	.2	.0004	.9003
7	.93	.1273	.0152	.9	.0015	.9013
8	1.26	.1126	.0218	1.3	.0022	[illegible]
9	1.58	.0979	.0211	1.2	.0021	.9061
10	1.91	.0698	.0179	1.1	[illegible]	.9073
11	2.24	.0518	.0141	.8	.0014	.9093
12	2.57	.0379	.0104	.6	[illegible]	.9103
13	2.90	.0275	[illegible]	[illegible]	[illegible]	[illegible]

PARAM. EST.DIST.LNORMAL

YMELN = 72.[illegible]7
XDELN = 123.79
COVLN = 1.72
SESLN = 10.22
CURLN = 404.6[illegible]

RESULTADOS DE LA PRUEBA X2(JI CUADRADO)

NIV CONF = 5.[illegible]

JI X2 TABULADO = 21.0[illegible]

DISTRIBUCION LOGNORMAL

INT	FRES OBS	FRES ESP	ERROR
1	2.000	******	******
2	2.000	18.667	14.882
3	9.000	11.642	3.738
4	5.000	6.256	.252
5	2.000	2.513	.1[illegible]5
6	5.000	.234	96.955
7	4.000	.893	10.776
8	9.000	1.285	46.325
9	8.000	1.245	36.640
10	8.000	1.058	45.571
11	5.000	.823	21.195
12	4.000	.615	18.635

JI X2 CALCULADO = *******

HO : CHIC = CHIT

H1 : CHIC MENORQUE CHIT

SI H1: SE ACEPTA EL AJUSTE COMO BUENO AL : 5.00DE CONFIANZA 1

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

1. PROBABILITY, STATISTICS, AND
   DECISION FOR CIVIL ENGINEERS
   JACK R. BENJAMIN
   HC GRAW HILL.

2. PROBABILITY AND STATISTICAL IN HIDROLOGY.
   VUJICA YEVJEVICH
   WATER RESOURCES PUBLICATIONS
   FORT COLLINS, COLORADO U.S.A.

3. SAMPLING TECHNIQUES
   WILLAM G. COCHRAN

4. HANDBOOK OF APLIED HIDROLOGY .
   (SECCION 8-1)
   VEN TE CHOW.

5. PROBABILIDAD Y ESTADISTICA
   SCHAWM.

6. APUNTES DE HIDROLOGIA
   SPRINGAL G. ROLANDO
   FACULTAD DE INGENIERIA

7. CALCULO
   HOWARD E. TAYLOR
   THOMAS E. WADE