

0835

IDENTIFICACION DE DISTRIBUCIONES DE VALORES EXTREMOS

JOSE FIDEL PEREZ DURAN

TRABAJO

PRESENTADO A LA DIVISION DE ESTUDIOS DE

POSGRADO DE LA

FACULTAD DE INGENIERIA

DE LA

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

COMO REQUISITO PARA OBTENER

EL GRADO DE

MAESTRO EN INGENIERIA

(APROVECHAMIENTOS HIDRAULICOS)

CIUDAD UNIVERSITARIA, D.F., 14 DE FEBRERO DE 1986



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

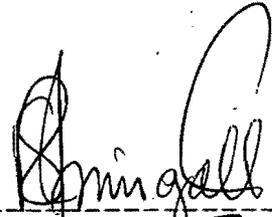
DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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.

APROBADO POR EL JURADO

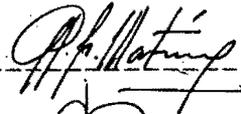
PRESIDENTE Dr. Rolando Springall Galindo



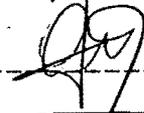
VOCAL Dr. Jose A. Raynal Villasenor



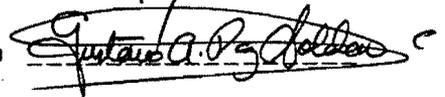
SECRETARIO Dr. Polioptro Martinez Austria



SUPLENTE M. en I. Oscar Fuentes Mariles



SUPLENTE M. en I. Gustavo A. Paz Soldan



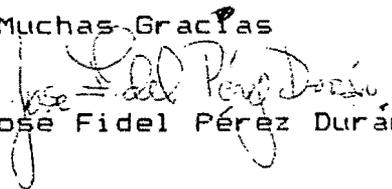
AGRADECIMIENTOS

Me considero bastante privilegiado de haber participado en la realización de un trabajo muy original e inédito. Las gracias por esto se las debo al Dr. José Angel Raynal Villasenor quien propuso y dirigió éste tema de investigación. Quedo además profundamente agradecido al Dr. Raynal por todo el apoyo y ayuda que me brindó: provisión de los artículos sobre los cuales se basa el trabajo; asesoramiento general; acceso a su clave en el sistema de computo "Cande" de esta universidad, e inclusive instrucciones y enseñanza sobre el mismo; acceso al procesador de palabras con el que se realizó este escrito; y todo el tiempo dedicado en cualquier instante que lo necesite. Agradezco también la gentileza mostrada por los compañeros Jaime Alfredo Flores Viveros, Daniel Francisco Campos Aranda y José Carlos Douriet Cardenas, ayudantes de el Dr. Raynal, mientras realice este trabajo, así como al personal de la unidad de computo de la DEPMI.

Con la División de Estudios de Posgrado de la Facultad de Ingeniería de esta Universidad Autónoma de México quedo en gran deuda; específicamente con las secciones de Aprovechamientos Hidraulicos y de Hidráulica. Le doy gracias a quienes me impartieron docencia y a todos quienes forman parte de la DEPMI.

No podría terminar esta parte sin dar gracias a quien a estado en todo lo que hay detras de esto y a quien estoy agradecido por haberme conducido y llevado hasta donde me encuentro:
DIOS

Muchas Gracias


José Fidel Pérez Durán

INDICE

RESUMEN

I.	INTRODUCCION	1
II.	PROPIEDADES Y CARACTERISTICAS DE LAS DISTRIBUCIONES DE VALORES EXTREMOS	5
III.	CRITERIOS DE IDENTIFICACION DE LAS DISTRIBUCIONES DE VALORES EXTREMOS	8
IV.	ANALISIS DE LOS CRITERIOS DE IDENTIFICACION A TRAVES DE MUESTREO DISTRIBUCIONAL	18
V.	ANALISIS Y DISCUSION DE RESULTADOS	20
VI.	CONCLUSIONES Y RECOMENDACIONES	30
VII.	REFERENCIAS	31
VIII.	LISTADO DEL PROGRAMA DE COMPUTADOR	33

RESUMEN

La distribución General de Valores Extremos esta compuesta por tres tipos que se distinguen por su parámetro de forma (Q). Estos tipos son: tipo I o Gumbel, cuando $Q=0$; tipo II o Fréchet, cuando $Q<0$; y tipo III o Weibull, cuando $Q>0$.

De varios criterios o pruebas que existen para la selección del tipo de distribución de valores extremos, tres de ellos fueron analizados a través de muestreo distribucional. Las pruebas son: la de Tiago de Oliveira; la de Hosking, Wood y Wallis; y la de Van Montfort. Para cada uno de los valores de el parámetro de forma -0.50 , -0.30 , -0.10 , 0.00 , 0.10 , 0.30 y 0.50 se generaron 5,000 muestras de tamaño $n=10$; 2,000 de $n=25$; y 1000 de $n=50$ con parámetros de ubicación y escala arbitrarios. A cada una de las 56,000 muestras generadas, y cuyo tipo de distribución está determinado por el valor de el parámetro de forma, se aplicaron los tres criterios para probar su eficiencia en identificar el tipo de distribución correcto.

Los resultados se consideran interesantes siendo que estas tres pruebas son de muy buena reputación en cuanto a su potencia se refiere. Ninguno de los tres criterios acerto sobre el tipo de distribución correcto en porcentajes deseablemente altos. Estos son : 27% el de Tiago de Oliveira; 17% el de Hosking, Wood y Wallis; y 45% el de Van Montfort. Los criterios muestran aumentar su eficiencia al aumentar el tamaño de las muestras. Son mucho menos eficientes al acercarse el valor del parámetro de forma a cero; especialmente por el lado negativo. Cuando la distribución es Gumbel los porcentajes de aciertos son de 95.54, 99.39 y 89.59 para el mismo orden de criterios de arriba. Sorprendentemente, el

critério de Tiago de Oliveira nunca detectó los casos en que la distribución era del tipo Fréchet o tipo II . La prueba de Hosking, Wood y Wallis es también de muy pobre eficiencia con este tipo II .Lo anterior muestra que los criterios tienen preferencia por y debilidades con algunos de los tipos de distribución. De los tres criterios analizados, el criterio de Van Montfort resultó ser superior en todos los aspectos. Se recomienda ,en clara superioridad sobre los criterios de Tiago de Oliveira y de Hosking, Wood y Wallis, el criterio de Van Montfort para $n < 50$ y valores del parámetro de forma entre -0.50 y 0.50 .

CAPITULO 1

INTRODUCCION

En los estudios hidrológicos, el análisis de los eventos extremos, como son las sequías y las grandes avenidas, tiene una gran importancia. La razón es que el dimensionamiento de las obras de aprovechamientos hidráulicos (su diseño) tiene que prestar atención especial a los requerimientos extremos para adecuar la obra o estructura y proveerle seguridad. En el caso de las obras de almacenamiento de agua la variable por analizar es el escurrimiento en los cauces.

La variable escurrimiento es aleatoria, lo cual implica que para su análisis se puede aplicar la probabilidad y la estadística, la cual encuentra su mayor aplicación en el análisis de frecuencias. El asunto de interés en este análisis es la excedencia de escurrimientos o la probabilidad de que el escurrimiento sea mayor que ciertos valores. Esto a su vez determina el período de retorno del evento. Aunque a priori no es posible determinar qué distribución representa mejor una muestra, existen formas o métodos aplicados con base en prueba y error, que permiten elegir la función de distribución apropiada. Afortunadamente, además, ciertos fenómenos y variables asociadas a ellos son reconocidas, por la experiencia, que están asociadas a cierta distribución o tipos de distribuciones. En el análisis de frecuencia este es el caso y dentro de las distribuciones alternativas se encuentra la Distribución General de Valores Extremos (GVE).

La distribución GVE está compuesta por, o reúne, tres tipos de distribuciones que son llamados: Gumbel o tipo I, Fréchet o tipo II, y Weibull o tipo III. La distribución GVE es una expresión general, precisamente, de estas tres. Interesante es saber que

la selección del tipo de distribución GVE es asunto delicado y no tan sencillo. En la actualidad no existe un medio teórico para deducir el tipo de distribución y la evidencia empírica apoya a más de un tipo de distribución; incluyendo a otros tipos de distribuciones de valores extremos no pertenecientes a la distribución GVE.

Todas las distribuciones tienen parámetros que las caracterizan y que en análisis hidrológicos son las bases de estudio. En el caso de la distribución GVE los parámetros son los tres siguientes: parámetro de ubicación, parámetro de escala y parámetro de forma. El comportamiento en las colas o extremos de las distribuciones está influenciado por uno de ellos solamente: el parámetro de forma. Para el tipo I este parámetro es igual a cero; para el tipo II, es menor que cero; y para el tipo III, es mayor que cero. El tipo de distribución GVE está determinado por el valor del parámetro de forma para una muestra dada.

Lo que se acaba de mencionar hace parecer que la selección del tipo de distribución GVE es relativamente fácil mediante la estimación de parámetros. El problema radica en que los métodos de estimación de parámetros no son del todo certeros y tienen probabilidades de error asociadas a su aplicación. Esta es la razón por la cual existen las pruebas de bondad de ajuste en las que, para un cierto nivel de confianza y basados en un indicador, se acepta o se rechaza la hipótesis de que una distribución representa la muestra. El problema se torna más interesante por el hecho de que las pruebas comunes de bondad de ajuste, como son la de Kolmogorov-Smirnov y la de Chi-cuadrado, no son apropiadas para hacer la distinción en el caso de las distribuciones de valores extremos. Los resultados de experimentos encontrados en N.E.R.C. (1975, ver pag. 135), en los que se aplican estas dos pruebas de bondad de ajuste, muestran que ninguna de ellas es de

mucha ayuda para identificar el tipo de distribución de valores extremos. La prueba de Chi-cuadrado rechaza diferentes distribuciones en números de veces similares y por lo tanto no apoya la selección de una distribución en particular. La prueba de Kolmogorov-Smirnov no parece ser sensitiva a desviaciones de las distribuciones de valores extremos. Por esta razón han surgido otros métodos o criterios para la identificación de distribuciones de valores extremos.

Hosking(1984) , cito y analizo trece criterios o pruebas para la selección del tipo de distribución GVE. Otten y Van Montfort (1978), añaden dos criterios más y Hosking, Wood y Wallis(1985) presentan otro más. En total existen , pues, 16 criterios que son: las dos pruebas de Wald; las dos pruebas "C(a)"; la prueba del multiplicador de Lagrange o prueba de "score"; las dos pruebas del cociente de verosimilitud; la prueba de Tiago de Oliveira; las tres pruebas de Stephen; la prueba de Bardsley; las dos pruebas de Gumbel; la prueba de Van Montfort; y la prueba de Hosking, Wood, y Wallis.

Algunos autores como Hosking(1984), se han preocupado por hacer comparaciones entre estos criterios. Otten y Van Montfort (1978), compararon el estadístico B de la prueba de Bardsley con el estadístico A propuesto por ellos mismos. En otro trabajo Van Montfort y Gomes(1985) compararon el criterio de Tiago de Oliveira, los dos estadísticos de Gumbel y de nuevo el estadístico A. En los dos últimos trabajos la comparación se baso en la potencia o función de potencia de la prueba. En el primero, lo que se hizo fue comparar las trece pruebas en terminos de su nivel de significancia empírico y la potencia en muestras pequeñas. De la tercera comparación citada, sus autores concluyeron que entre los estadísticos G, G^* , T (de Tiago), y el A, no hay diferencia de potencia significativa; el G es preferible sobre el T por ser más simple; y el A sobre el G dependiendo de la disponibilidad de

instrumentos de computo. En el segundo trabajo, sobre los estadísticos A y B, se concluyó que A es ligeramente más poderoso que B. De la comparación de las trece pruebas por Hosking, la del cociente de verosimilitud resulto ser la prueba más poderosa y luego la de Van Montfort que Hosking recomienda.

Es oportuno mencionar que la prueba de Hosking, Wood y Wallis no ha sido comparada con otras; probablemente por ser de reciente proposición. Otro detalle de importancia, y que es necesario mencionar, es que las comparaciones con base en la potencia de las pruebas no son muy concluyentes en cuanto a la eficiencia de las mismas como más adelante se comprobara.

Esta investigación ha tenido por propósito analizar tres de los criterios de distinción de distribución de valores extremos. Las pruebas elegidas son la de Van Montfort, la de Tiago de Oliveira y la de Hosking, Wood y Wallis. La comparación se hace con base en un experimento sencillo que es el de aplicar los criterios a muestras cuyo tipo de distribución es conocido y ver que tan buenos son los criterios en detectar la correcta. Esto, que no ha sido realizado para estos criterios, es mucho más práctico y concluyente que las comparaciones hechas anteriormente y mencionadas arriba.

Antes de pasar al análisis de los criterios, se hace necesario presentar una descripción breve de la distribución GVE y sus tres tipos, así como de los tres criterios elegidos. Esto se hace en los capítulos 2 y 3. En el capítulo 4 se detalla el análisis de los criterios, presentando los resultados y discutiéndolos en el capítulo 5. Finalmente, en el capítulo 6 se presentan las conclusiones y recomendaciones.

CAPITULO 2
 PROPIEDADES Y CARACTERISTICAS DE LAS
 DISTRIBUCIONES DE VALORES EXTREMOS

El valor extremo de un conjunto de variables aleatorias es también una variable aleatoria. La distribución GVE y sus tres tipos tienen la habilidad de modelar muy bien el comportamiento de estas variables de valores extremos; de ahí su nombre. La distribución GVE esta compuesta por tres tipos denominados I, II y III. Estas distribuciones se relacionan con las distribuciones de donde provienen los valores extremos, o distribuciones "padre" (ver Haan(1977), pp.110-119); en muchos casos la distribución padre no es conocida. Las propiedades y características de los tres tipos se presentan a continuación en la tabla 2.1. Estas propiedades se pueden encontrar en varias fuentes. Las expresiones presentadas aquí (sin derivación alguna) se encuentran en Raynal (1985).

La función de distribución de probabilidades de la distribución GVE y su función de densidad de probabilidades son:

$$F(x) = \exp \left[- \left[1 - \left(\frac{x-\mu}{\sigma} \right) Q \right]^{1/Q} \right] \quad (2.1)$$

$$f(x) = \frac{1}{\sigma} \left[1 - \left(\frac{x-\mu}{\sigma} \right) Q \right]^{1/Q-1} F(x) \quad (2.2)$$

TIPO I: GUMBEL

La función de distribución GVE evaluada en $Q \rightarrow 0$ resulta ser la función de distribución tipo I o Gumbel. Se le conoce también

como Fisher-Tippet I y doble exponencial. Es la más conocida de las tres y la de mayor aplicación. Es una distribución simétrica y sin restricciones a la derecha y a la izquierda. Esta característica la tienen muchas variables hidrológicas. Los logaritmos de variables aleatorias que siguen el tipo III tienen distribución tipo I. Para valores máximos se relaciona con las distribuciones normal, lognormal, exponencial y gamma. Se puede aplicar a mínimos de la normal.

TIPO II: FRECHET

Esta es la función en la que el parámetro de forma es menor que cero ($Q < 0$). Es la menos conocida de las tres y no ha tenido mucha aplicación. Se puede aplicar para máximos y mínimos de extremos provenientes de distribuciones del tipo Cauchy.

TIPO III: WEIBULL

Esta es la distribución en la que $Q > 0$. El mayor uso de esta distribución es como distribución de escurrimientos mínimos o bajos, los cuales están limitados por la izquierda por cero. Está relacionada con las distribuciones beta, lognormal, gamma y exponencial. Para máximos es posible aplicarla a la beta.

TABLA 2.1

DISTRIBUCION	TIPO I. GUMBEL	TIPO II FRECHET	TIPO III WEIBULL
condiciones	$y = \frac{(x-\mu)}{\sigma}Q \quad -\infty < y < +\infty$ $Q=0 ; \sigma > 0 ; -\infty < \mu < \infty$	$y = 1 - \left(\frac{x-\mu}{\sigma}\right)Q \quad 0 < y < \infty$ $\mu + \sigma/Q \leq x < \infty$ $Q < 0 ; \sigma > 0 ; \gamma \leq 1.14$	$-y = 1 - \left(\frac{x-\mu}{\sigma}\right)Q \quad -\infty < y \leq 0$ $-\infty < x \leq \mu + \sigma/Q$ $Q > 0 ; \sigma > 0 ; \gamma < 1.14$
moda	$mo = \mu$	$mo = (1+Q)^Q$	$mo = -(1-Q)^Q$
mediana	$med = \mu$	$med = (\ln 2)^Q$	$med = -(\ln 2)^Q$
media	$E = \mu \pm 0.577\sigma$	$E = \Gamma(1+Q)$	$E = -\Gamma(1+Q)$
varianza	$Var = 1.645\sigma^2$	$Var = \Gamma(1+2Q) - \Gamma^2(1+Q)$	$Var = \Gamma(1+2Q) - \Gamma^2(1+Q)$
coeficiente de variacion	$CV = \left(0.7797\frac{\mu}{\sigma} + 0.45\right)^{-1}$	$CV = \left[\frac{\Gamma(1+2Q)}{\Gamma^2(1+Q)} - 1\right]^{1/2}$	$CV = \left[\frac{\Gamma(1+2Q)}{\Gamma^2(1+Q)} - 1\right]^{1/2}$
coeficiente de asimetria	$\gamma = \pm 1.1396$	$\gamma = \frac{\Gamma(1+3Q) - 3\Gamma(1+2Q)\Gamma(1+Q) + 2\Gamma^3(1+Q)}{(\Gamma(1+2Q) - \Gamma^2(1+Q))^{1.5}}$	$\gamma = \frac{-\Gamma(1+3Q) + 3\Gamma(1+2Q)\Gamma(1+Q) - 2\Gamma^3(1+Q)}{(\Gamma(1+2Q) - \Gamma^2(1+Q))^{1.5}}$
fn de distribucion	$F = \exp(-\exp(\pm y))$	$F = \exp(-y^{1/Q})$	$F = \exp((-y)^{1/Q})$
fn de densidad	$f = \exp(\pm y - \exp(\pm y))$	$f = -\frac{y^{1/Q-1}}{Q} F(y)$	$f = \frac{1}{Q} (-y)^{1/Q-1} F(y)$

parametros: μ de ubicacion; σ de escala; y Q de forma

CAPITULO 3

CRITERIOS DE IDENTIFICACION DE LAS DISTRIBUCIONES DE VALORES EXTREMOS

El propósito de este capítulo es describir los criterios de identificación de el tipo de distribución de valores extremos seleccionados para este estudio. Estos son los criterios de Tiago de Oliveira, Van Montfort y el de Hosking, Wood y Wallis. No se pretende más que describirlos en cuanto a su aplicación se refiere, que es el objetivo específico de este trabajo. Antes de iniciar el tema, algo debe mencionarse acerca de las pruebas estadísticas, su procedimiento y características generales.

En la estadística es muy común el problema de que teniendo una muestra $x(1), x(2), \dots, x(n)$, de una variable aleatoria X , interesa saber si pertenecen a una distribución $f(x)$, o a otra $g(x)$. Esto es precisamente lo que se busca determinar en la distribución GVE y sus tres tipos. Para demostrar una cosa o la otra, se debe aplicar alguna prueba y la prueba debe basarse en una hipótesis, ya que sin esta el problema permanece sin ser atacado (la muestra pertenece a cualquiera de las distribuciones). Una hipótesis estadística, denotada por H_0 , es una aseveración o conjetura acerca de la distribución de una variable aleatoria (ver Mood et al (1974) p.402). Una vez definida la hipótesis, debe procederse a probar si es cierta o falsa. El procedimiento general de una prueba estadística (ver Haan (1977) p.166) es el siguiente:

1. formular la hipótesis a ser probada
2. formular la hipótesis alternativa
3. determinar un estadístico de prueba
4. determinar la distribución de el estadístico
5. definir la región de rechazo (crítica) del estadístico

6. calcular el valor del estadístico para una muestra
7. comparar este valor con los de la región crítica y determinar si está dentro o fuera de ella

Los pasos 1 al 5 constituyen la formulación de la prueba. Los pasos 6 y 7 constituyen la realización de la prueba. La hipótesis a ser probada se denomina hipótesis nula. La forma de definir esta hipótesis, y la prueba, depende de el problema de interés. Muchas pruebas se basan sobre los parámetros de la distribuciones. En este caso es una buena opción basar la prueba sobre el parámetro de forma de la distribución GVE. La regla de decisión no tiene que ser este parámetro directamente, pero el estadístico debe comportarse de manera diferente bajo las hipótesis consideradas y conforme a las características del parámetro sobre el cual se basa la distinción. Un buen candidato para ser estadístico de prueba es un buen estimador del parámetro que es utilizado para especificar la hipótesis o un estadístico relacionado a éste.

Los criterios analizados tienen como hipótesis nula $H_0: Q=0$, donde Q es el parámetro de forma de la distribución GVE. La hipótesis alternativa es $H_1: Q$ es diferente de cero. En este caso la prueba se considera de doble alternativa o de dos lados ya que las alternativas son $Q < 0$ y $Q > 0$. Todos los estadísticos que se propongan deben ser sensibles al comportamiento de este parámetro solamente.

Un buen indicador de que tan buena es una prueba es la potencia de la prueba. La potencia de la prueba se relaciona con los posibles errores de la prueba. Estos errores pueden ser dos: rechazar H_0 cuando es verdadera y aceptarla cuando es falsa. La función de potencia de la prueba se define como la probabilidad de rechazar H_0 cuando es una hipótesis verdadera (error tipo I). La función de potencia ideal de una prueba es una que es igual a cero para el parámetro de la hipótesis nula e igual a la unidad

para la hipótesis alternativa. Es con base en la potencia de la prueba que se han basado las comparaciones de los criterios de identificación de las distribuciones de valores extremos. Se presentan a continuación los tres aquí analizados. Al final del capítulo se incluyen tres ejemplos en los que se muestra la aplicación de los tres criterios seleccionados.

CRITERIO DE TIAGO DE OLIVEIRA

Tiago de Oliveira (1981) desarrolló una prueba poderosa que es del tipo de las del multiplicador de Lagrange. La regla de decisión de este criterio es la siguiente:

- si $V_n < T$ elíjase la tipo III (Weibull)
- si $\text{abs}(V_n) < T$ elíjase la tipo I (Gumbel)
- si $V_n > T$ elíjase la tipo II (Fréchet)

La prueba se puede llevar a cabo con los siguientes elementos:

media de la muestra de las X's: X_m

desviación estandar: S_x

$$\text{parámetro de ubicación: } \mu = X_m - 0.455S_x \quad (3.1)$$

$$\text{parámetro de escala: } \sigma = 0.7806S_x \quad (3.2)$$

$$\text{variable estandarizada: } Z_i = (X_i - \mu) / \sigma \quad (3.3)$$

estadístico de prueba: V_i

$$V_i = -Z_i + [1 - \exp(-Z_i)](Z_i)^2 / 2 \quad (3.4)$$

V_n : media aritmética de V_i

$$T = (2.09797/n)^{0.5} C_n \quad (3.5)$$

C_n es una variable con distribución normal

Van Montfort y Gomes (1985) compararon este criterio con otros dos. La prueba resulto tener una potencia similar a las de Gumbel y la de Van Montfort. Ellos la consideran de aplicación complica-

da para preferirla en lugar de las otras dos. Tiago de Oliveira (1981) dice que la prueba parece ser util solo para tamaños de muestras mayores de 400.

CRITERIO DE VAN MONTFORT

Originalmente Van Montfort (1970) desarrolló o encontró un estadístico de prueba " $Z=f(r)$ ". Mas tarde Van Montfort y Otten (1978) modificaron el estadístico y lo presentaron en la forma que se presenta aquí. Los elementos requeridos para la aplicación de el criterio de Van Montfort, para una muestra de valores ordenados de X , son:

$$\Delta_i = \text{Ln}[-\text{Ln}\{(i-.5)/(n+1)\}] \quad ; i=2, \dots, n \quad (3.6)$$

$$\Delta_m = \sum(\Delta_i) / (n-1) \quad (3.7)$$

$$\sigma_{\Delta}^2 = \sum(\Delta_i - \Delta_m)^2 / (n-1) \quad (3.8)$$

$$M_i = -\text{Ln}[-\text{Ln}(i/(n+1))] \quad ; i=1, \dots, n \quad (3.9)$$

$$L_i = (X(i) - X(i-1)) / (M(i) - M(i-1)); \quad i=2, \dots, n \quad (3.10)$$

$$A = [\sum(L_i \Delta_i) / \sum(\Delta_i) - \Delta_m] / (\sigma_{\Delta}^2 / n)^{0.5} \quad (3.11)$$

Esta prueba con el estadístico A resultó ser un poco más poderosa que la de Bardsley segun Otten y Van Montfort (1978). En la comparación de Hosking (1984), resultó estar sólo por debajo de el criterio o prueba de el cociente de verosimilitud y Hosking la recomienda.

Para encontrar los valores críticos de A , se simuló la distribución de A . Para cada tamaño de muestra se generaron 1999 números aleatorios que fueron transformados en nueve variables cada uno con $Q = -.8, -.4, -.2, -.1, 0.0, .1, .2, .4, \text{ y } .8$. Se calculó el valor de A para cada muestra y de la función de distribución simulada se leyeron los valores críticos de A en $Q=0$. Los valores se muestran en la tabla 3.1, tomada de Otten y Van Montfort (1978).

La prueba consiste en calcular el valor de A para la muestra dada y compararlo con el valor crítico de A para el tamaño de muestra y nivel de significancia correspondientes. Si A resulta ser positivo, la alternativa posible es que $Q > 0$. Si el A calculado es menor o igual que el de la tabla, se acepta la hipótesis de $H_0: Q = 0$ y la distribución es del tipo I o Gumbel; en caso contrario es del tipo III o Weibull. Cuando el A calculado es negativo, la alternativa posible es que $Q < 0$. Si el valor de A calculado es mayor o igual que el de la tabla, se acepta la hipótesis de $Q = 0$ y la distribución es del tipo I; en caso contrario la distribución es tipo II o Fréchet.

CRITERIO DE HOSKING, WOOD Y WALLIS

Hosking, Wood y Wallis (1985) propusieron la prueba más reciente y está basada en la estimación de parámetros por el método de momentos de probabilidad pesada (MFP) para el cual ellos proponen una variante. Su método de MFP se calcula para una muestra ordenada de X, de la siguiente manera:

$$B_0 = \sum (X(i)) / n \quad (3.12)$$

$$B_1 = \sum (X(i) \cdot (i-1)) / n \cdot (n-1) \quad (3.13)$$

$$B_2 = \sum (X(i) \cdot (i-1) \cdot (i-2)) / n \cdot (n-1) \cdot (n-2) \quad (3.14)$$

$$C = \{(2 \cdot B_1 - B_0) / (3 \cdot B_2 - B_0)\} - \{\ln(2) / \ln(3)\} \quad (3.15)$$

$$Q_{est.} = 7.8590 \cdot C + 2.9554 \cdot C^2 \quad (3.16)$$

El estadístico que estos autores han propuesto es $Z = (n/0.56633)^{0.6} Q_{est.}$. La prueba consiste en calcular el estimado de el parámetro de forma $Q_{est.}$, calcular Z y compararlo con el valor crítico de Z. Los valores críticos de Z están en la tabla 3.2, tabla que está basada en simulación vía computadora de 50,000 muestras para cada valor de n (tamaño de muestra) y $Q_{est.}$, y se encuentra en Hosking, Wood y Wallis (1985).

Según sus autores la prueba tiene una potencia similar a la del cociente de verosimilitud que Hosking(1984) había encontrado ser la más poderosa de las trece pruebas que comparó. Es una prueba sencilla, poderosa y precisa.

TABLA 3.1

Valores críticos de el estadístico A de Van Montfort.

-DIST	$Q < 0$			$Q > 0$		
	0.025	0.05	0.10	0.10	0.05	0.025
n						
10	-2.6	-1.72	-1.37	1.17	1.43	1.68
25	-2.4	-1.83	-1.45	1.10	1.39	1.60
50	-2.35	-1.78	-1.46	1.02	1.37	1.60
100	-2.27	-1.89	-1.46	1.03	1.31	1.57

TABLA 3.2

Niveles de significancia empíricos para el estadístico Z y la prueba con hipótesis $H_0: Q=0$ contra las alternativas $Q < 0$ y $Q > 0$.

n	$Q < 0$		$Q > 0$	
	10%	5%	10%	5%
15	10.3	4.3	7.3	3.7
25	10.4	4.6	8.4	4.3
50	10.5	4.9	8.9	4.6
100	10.4	5.1	9.4	4.9
200	10.4	5.0	9.7	5.1
500	10.5	5.3	9.6	4.9

EJEMPLPO 3.1 :Criterio de Tiago de Oliveira

i	Xi	Zi	Vi
1	19.885	-1.6262	-3.7741
2	20.940	-1.1201	-0.1755
3	21.820	-0.6981	0.4520
4	23.700	0.2037	-0.1998
5	24.888	0.7735	-0.6124
6	25.460	1.0478	-0.6914
7	25.760	1.1917	-0.6973
8	26.720	1.6522	-0.5489
9	27.500	2.0263	-0.2440
10	28.100	2.3141	0.0987
To	244.773		-6.3927

$X_{media} = 24.4773$

$S_x = 2.6709$

$\mu = 23.2754$

$\sigma = 2.0849$

$V_n = -0.6393$

$T = 0.8931$

-como $T > V_n$, no es tipo II ($T < V_n$)

-como $-T > V_n$ y $T > V_n$, no es tipo I ($-T < V_n < T$)

-como $-T > V_n$, si es tipo III ($-T > V_n$)

EJEMPLO 3.2: Criterio de Van Montfort

i	X_i	Δ_i	M_i	$X_0 - X_i$	$M_0 - M_i$	L_i	$\Delta_i \cdot L_i$
1	19.885	-	-0.8746	-	-	-	-
2	20.940	0.6894	-0.5334	1.055	0.3412	3.0923	2.1317
3	21.820	0.3931	-0.2618	0.880	0.2716	3.2400	1.2737
4	23.700	0.1355	-0.0115	1.880	0.2503	7.5116	1.0180
5	24.888	-0.1123	0.2377	1.188	0.2492	4.6770	-0.5351
6	25.460	-0.3665	0.5007	0.572	0.2630	2.1751	-0.7972
7	25.760	-0.6423	0.7941	0.300	0.2935	1.0223	-0.6566
8	26.720	-0.9597	1.1443	0.960	0.3502	2.7415	-2.6311
9	27.500	-1.3555	1.6061	0.780	0.4618	1.6890	-2.2894
10	28.100	-1.9200	2.3506	0.600	0.7445	0.8059	-1.5473
To.		-4.1383				27.0448	-4.0334

$$\Delta_m = -0.4598$$

$$\sigma_{\Delta}^2 = 0.6388$$

$$A = 1.2291$$

El valor crítico de A (para $n=10$ y $\text{sign.}=0.05$; y $Q>0$ ya que $A>0$) es 1.43 (ver tabla 3.1)

-como $A_{\text{calculado}} < A_{\text{tabulado}}$, la hipótesis $H_0: Q=0$ se acepta y la distribución es del tipo I o Gumbel

EJEMPLO 3.3: Criterio de Hosking, Wood y Wallis

i	X(i)	X*(i-1)	X*(i-1)*(i-2)
1	19.885	-	-
2	20.940	20.940	-
3	21.820	43.640	43.640
4	23.700	71.110	142.200
5	24.800	99.200	297.600
6	25.460	127.300	509.200
7	25.760	154.560	772.800
8	26.720	187.040	1,122.240
9	27.500	220.000	1,540.000
10	28.100	252.900	2,023.200
11	28.600	286.000	2,574.000
12	30.200	332.200	3,322.000
13	30.380	364.560	4,010.160
14	31.500	409.500	4,914.000
15	32.600	456.400	5,933.200
To	397.965	3025.340	27,204.240

$B_0=26.5310$; $B_1=14.4064$; $B_2=9.9649$

$C=0.0474$; $Q_{est.}=0.3792$; $Z=1.9568$

$Z_{tab}(\text{para } n=15, \text{sign.}=0.05 \text{ y } Q>0)=3.70$

-como $Z < Z_{tab}$, la distribución es del tipo I o Gumbel

CAPITULO 4

ANALISIS DE LOS CRITERIOS DE IDENTIFICACION A TRAVES DE MUESTREO DISTRIBUCIONAL

El análisis de los criterios de identificación de distribuciones GVE planteado y realizado consiste en un experimento sencillo pero diferente a lo realizado sobre este tema anteriormente. Como se mencionó, las comparaciones entre los criterios se han basado en la función de potencia de los estadísticos correspondientes. Lo que se hizo aquí es algo mucho más práctico y es poner a prueba los tres criterios elegidos, aplicandolos a muestras cuyos tipos de distribuciones se fijaron de antemano (de aquí lo de muestreo distribucional). El criterio que acierte el mayor número de veces ha de ser el más eficiente (para los casos analizados) independientemente de su potencia.

Para tener una muestra con una distribución deseada, se aplica o se hace uso de el teorema de "la transformada integral de probabilidad". Este teorema dice (ver Mood et al (1974)p.202): "Si X es una variable aleatoria con función acumulativa de distribución continua $F(x)$, entonces $U=F(x)$ se distribuye uniformemente sobre el intervalo $(0,1)$. De manera inversa, si U se distribuye uniformemente en el intervalo $(0,1)$, entonces $X=F^*(U)$ tiene una función acumulativa de distribución de probabilidades $F(\cdot)$, donde F^* es la función inversa de $F(x)$ ". La generación de datos es relativamente sencilla. El procedimiento es: (1) generar un número aleatorio U de la función de distribución uniforme (o rectangular); (2) igualar este número a $F(x)$; y (3) encontrar el valor x con la función inversa de x , p.e. F^* . Los valores de x así generados estarán representados por la función $F(x)$ elegida, y esa muestra de la variable aleatoria X por los parámetros empleados para definir la función. Todo esto es sólo posible si se

conoce la función inversa de la distribución de interés , caso que afortunadamente es el de la distribución GVE. Esta función inversa es:

$$x = -\text{Ln}(-\text{Ln}(F(x)))\sigma + \mu \quad \text{para } Q=0 \quad (4.1)$$

$$y \quad x = \mu + \sigma(1 - (-\text{Ln}(F(x)))^Q)/Q \quad \text{para } Q \neq 0 \quad (4.2)$$

Para realizar el experimento se crearon programas de computadora que aplican los métodos que se describen en el capítulo 3, y luego se integraron en uno solo. Se fijó un nivel de confianza de 95%. Los valores del parámetro de forma seleccionados son: - 0.50, -0.30, -0.10, 0.0, 0.10, 0.30 y 0.50. Estos valores cubren el rango en el cual se presenta el parámetro de forma en la práctica. Los números aleatorios se tomaron de un archivo que disponía de un número suficiente de estos para los fines propuestos. Se eligieron tamaños de muestras de 10, 25 y 50, considerando de mayor interés analizar muestras de tamaños realistas. Se generaron 5000 muestras de tamaño $n=10$; 2000 de $n=25$; y 1000 de $n=50$. En cada uno de los tres casos se aplicó la función inversa para cada uno de los siete valores del parámetro de forma, resultando ser analizadas un total de 56,000 muestras. Para el caso de $Q=0$, se aplica la primera de las funciones inversas de arriba(4.1).

Ya que los criterios de identificación tienen la característica de ser invariantes con respecto a los parámetros de ubicación y escala, los valores de estos no afectan la aplicación de las pruebas ni la calidad de los resultados. Los valores empleados para el parámetro de ubicación y el de escala son de 10.0 y 4.0 respectivamente. Los resultados se presentan y discuten en el siguiente capítulo.

CAPITULO 5

ANALISIS Y DISCUSION DE LOS RESULTADOS

Se analizaron un total de 56,000 muestras; 8,000 muestras para cada uno de los 7 parámetros de forma. De estas, 24,000 se generaron con distribución Fréchet (tipo II); 8,000 con Gumbel (tipo I); y 24,000 con Weibull (tipo III). Las cantidades de muestras para los diferentes tamaños fueron: 35,000 para $n=10$; 14,000 para $n=25$; y 7,000 para $n=50$. Los resultados del análisis se presentan en este capítulo. Ya que una tabla puede decir más que mil palabras, se decidió presentar varias tablas que analizan o muestran además de los resultados de índole cuantitativo, aquellos de índole cualitativo y que se comentan y destacan a continuación.

En cuanto a totales se refiere, los criterios analizados no muestran ser muy efectivos. El criterio de Van Montfort acertó en un 45% de los casos, siendo seguido por el de Tiago de Oliveira con 29%, y el de Hosking, Wood y Wallis lo hizo en apenas un 17%. Los porcentajes se refieren a la cifra de 56,000 muestras que se analizaron.

Un factor que requiere mención es el efecto del tamaño de las muestras. En los tres criterios, el porcentaje de aciertos aumenta desde para $n=10$ hasta para $n=50$. El aumento es para los criterios de Tiago de Oliveira y el de Van Montfort de un 100% aproximadamente (ver tabla 5.8).

El análisis revela un hecho importante y que es el reflejado en las cifras totales de arriba: los criterios tienen debilidades en y/o preferencia por algún tipo de distribución. En los 8,000 casos en que la distribución fue Gumbel, los criterios resultan ser bastante buenos con porcentajes de aciertos de 96, 99 y 90

para los criterios de Tiago de Oliveira, el de Van Montfort y el de Hosking, Wood y Wallis respectivamente. Para los casos en que la distribución fue Fréchet (24,000), los criterios o dejan que desear o sencillamente no detectan este tipo de distribución. El criterio de Tiago de Oliveira no detectó ninguna de las 24,000 muestras con distribución Fréchet como tales. El de Hosking, Wood y Wallis lo hizo solamente en 2% de estos casos, mientras que el de Van Montfort en un 39% de estos. Para la distribución Weibull las cifras no son tan bajas pero tampoco indican mucha efectividad. Estas cifras son en el mismo orden de criterios: 35%, 10% y 36% . . .

El efecto del valor de el parámetro de forma también se deja notar. El criterio de Tiago de Oliveira, que no detectó la distribución Fréchet, detectó la distribución Weibull para los casos $Q = .10, .30$ y $.50$ en porcentajes de 12, 36 y 56 respectivamente. Para el criterio de Hosking, Wood y Wallis con esta misma distribución las cifras van desde el 1% de aciertos hasta el 23%. Este criterio muestra un resultado similar para la distribución Fréchet, siendo de 0% la cantidad de aciertos cuando $Q = -.10$. En el criterio de Van Montfort se nota lo mismo pero las cifras son más elevadas en valor numérico, y mostrando ser muy superior en el caso de la distribución Fréchet. Los tres criterios tienen, pues, mayor dificultad para detectar la distribución correcta al acercarse el valor de el parámetro de forma a cero (ver tabla 5.9).

Los dos efectos anteriores se ven a su vez afectados por el tamaño de las muestras. O sea, las variaciones en los porcentajes de acierto que se notan con los tipos de distribución y con los distintos parámetros de forma son diferentes para los tamaños de muestra de 10, 25 y 50 . Esto quiere decir que el análisis y discusión presentado requiere una consideración bidimensional, razón por la cual están las tablas.

Las tablas 5.1 a 5.7 muestran el número de muestras que cada criterio identificó como correspondiente a las distribuciones.

TABLA 5.1

Criterio	Tiago	Hosking	Van Montfort
Q=-0.50			
n=10 Gum	4990	4940	2675
Fre	0	57	2311
Wei	10	3	14
n=25 Gum	2000	1863	416
Fre	0	137	1584
Wei	0	0	0
n=50 Gum	1000	720	43
Fre	0	280	957
Wei	0	0	0

TABLA 5.2

Criterio	Tiago	Hosking	Van Montfort
Q=-0.30			
n=10 Gum	4957	4985	3539
Fre	0	5	1414
Wei	43	10	47
n=25 Gum	2000	1983	911
Fre	0	17	1088
Wei	0	0	1
n=50 Gum	1000	956	236
Fre	0	44	764
Wei	0	0	0

TABLA 5.3

Criterio Tiago Hosking Van Montfort
 Q=-0.10

n=10	Gum	4852	4974	4272
	Fre	0	0	586
	Wei	148	26	142
n=25	Gum	1956	2000	1603
	Fre	0	0	358
	Wei	44	0	39
n=50	Gum	991	1000	729
	Fre	0	0	262
	Wei	9	0	9

TABLA 5.4

CRITERIO Tiago Hosking Van Montfort
 Q=0

n=10	Gum	4765	4951	4478
	Fre	0	0	282
	Wei	235	49	240
n=25	Gum	1887	2000	1794
	Fre	0	0	108
	Wei	113	0	98
n=50	Gum	948	1000	895
	Fre	0	0	53
	Wei	52	0	52

TABLA 5.5

Criterio Tiago Hosking Van Montfort
 Q=0.10

n=10	Gum	4594	4912	4473
	Fre	0	0	125
	Wei	406	88	402
n=25	Gum	1708	1999	1736
	Fre	0	0	18
	Wei	292	1	246
n=50	Gum	768	999	794
	Fre	0	0	2
	Wei	232	1	204

TABLA 5.6

Criterio Tiago Hosking Van Montfort
 Q=0.30

n=10	Gum	3993	4689	3972
	Fre	0	0	11
	Wei	1007	311	1017
n=25	Gum	924	1950	938
	Fre	0	0	0
	Wei	1076	50	1062
n=50	Gum	168	949	135
	Fre	0	0	0
	Wei	832	51	865

TABLA 5.7

CRITERIO Tiago Hosking Van Montfort
Q=0.50

n=10	Gum	3119	4131	2937
	Fre	0	0	2
	Wei	1881	869	2061
n=25	Gum	340	1583	251
	Fre	0	0	0
	Wei	1660	417	1749
n=50	Gum	17	468	5
	Fre	0	0	0
	Wei	983	532	995

TABLA 5.8

Número de aciertos por tipo de distribución y por tamaño de muestra (porcentajes entre paréntesis)

n=10 (35,000 muestras)

DIST. Fréchet (15,000) Gumbel (5,000) Weibul (15,000)
Cri.

Tiago	(0)	4765	(95.30)	3294	(21.96)
Hosking	62 (0.41)	4951	(99.02)	1268	(8.45)
VanMon	4311 (28.74)	4478	(89.56)	3480	(23.20)

n=25 (14,000 muestras)

Dist. Fréchet (6,000) Gumbel (2000) Weibull (6,000)
Cri.

Tiago	(0)	1887	(94.35)	3028	(50.47)
Hosking	154 (2.57)	2000	(100.0)	468	(7.80)
VanMon	3030 (50.50)	1794	(89.70)	3057	(50.95)

n=50 (7,000 muestras)

Dist. Fréchet (3,000) Gumbel (1,000) Weibull (3,000)
Cri.

Tiago	(0)	991	(99.10)	2047	(68.23)
Hosking	324 (10.80)	1000	(100.0)	584	(19.47)
VanMon	1983 (66.10)	895	(89.50)	2064	(68.80)

n=10,25,50 (56,000 muestras)

Dist. Fréchet (24,000) Gumbel (8,000) Weibull (24,000)
Cri.

Tiago		(0)	7643	(95.54)	8369	(34.87)
Hosking	540	(2.25)	7951	(99.39)	2320	(9.67)
VanMon	9324	(38.85)	7167	(89.59)	8601	(35.84)

TABLA 5.9

Números de aciertos de cada criterio en función de el parámetro de forma y el tamaño de las muestras

n=10 (35,000 muestras)

Dist.	Fréchet (15000)			Gumbel (5000)	Weibul (15000)		
p. frm	-0.50	-0.30	-0.10	0.00	0.10	0.30	0.50
Tiago	0	0	0	4765	406	1007	1881
Hosking	57	5	0	4951	88	311	869
VanMon	2311	1414	586	4478	402	1017	2061

n=25 (14,000 muestras)

Dist.	Fréchet (6,000)			Gumbel (2000)	Weibull (6000)		
p. frm	-0.50	-0.30	-0.10	0.00	0.10	0.30	0.50
Tiago	0	0	0	1887	292	1076	1660
Hosking	137	17	0	2000	1	50	417
VanMon	1584	1088	358	1794	246	1062	1749

n=50 (7,000 muestras)

Dist.	Fréchet (3,000)			Gumbel (1000)	Weibull (3000)		
p. frm	-0.50	-0.30	-0.10	0.00	0.10	0.30	0.50
Tiago	0	0	0	948	232	832	983
Hosking	280	44	0	1000	1	51	532
VanMon	957	764	262	895	204	865	995

n=10,25 y 50 (56,000 muestras)

Dist.	Fréchet (24000)			Gumbel (8000)	Weibull (24000)		
	p. frm -0.50	-0.30	-0.10	0.00	0.10	0.30	0.50
Tiago	0	0	0	7600	930	2915	4524
Hosking	474	66	0	7951	90	412	1818
VanMon	4852	3266	1206	7167	852	2944	4805



DEPTO.

CAPITULO 6

CONCLUSIONES Y RECOMENDACIONES

Ninguno de los tres criterios es muy eficiente en detectar el tipo correcto de distribución de valores extremos para tamaños de muestra menores de 50. Los tres criterios son mucho más efectivos cuando la distribución es verdaderamente de tipo I o Gumbel, mostrando debilidad en detectar los otros dos tipos. Al aumentar el tamaño de las muestras, los criterios mejoran su eficacia. Mientras más cercano de cero es el valor del parámetro de forma, se hace más difícil que cualquiera de estos tres criterios detecte la distribución correcta. Cuando la distribución es de tipo Fréchet los criterios de Tiago de Oliveira y de Hosking, Wood y Wallis son totalmente ineficientes.

En todos los aspectos mencionados existe superioridad de uno de los tres criterios. Esto facilita la aplicación de una escala de decisión sobre la selección de uno de los tres criterios. El criterio más efectivo para tamaños de muestras menores de 50 es el de Van Montfort. El criterio más efectivo en el rango $-0.50 < Q < 0.50$ es el de Van Montfort. Un criterio es muy superior a los otros dos en detectar el tipo de distribución cuando ésta es Fréchet y éste es el criterio de Van Montfort. Desde cualquier ángulo que se mire el mejor criterio es el de Van Montfort.

De los tres criterios analizados se recomienda, para $n < 50$ y $-0.50 < Q < 0.50$, el criterio de Van Montfort. Se reconoce, sin embargo que su eficacia no es lo deseablemente alta. No se recomienda el criterio de Tiago de Oliveira, o más bien se sugiere que se descarte para los rangos de tamaño de muestra y de parámetro de forma considerados. Lo mismo se hace con el criterio de Hosking, Wood y Wallis.

REFERENCIAS

1. Haan, Charles T., "Statistical methods in hydrology", The Iowa State University Press, 1977. 378 p.
2. Hosking, J.R.M. 1984, "Testing whether the shape parameter is zero in the generalized extreme-value distribution". Biometrika 71-2, pp.367-374
3. Hosking, J.R.M., Wallis, J.R. y Wood, E.F. (1985). "Estimation of the generalized extreme-value distribution by the method of probabilities weighted moments". Technometrics 27-3; pp. 251-262
4. Mood, A.M., Graybill, F.A. and Boes, D.C. "Introduction to the theory of statistics", 3rd ed. McGraw Hill, Auckland, 1974. 564 p.
5. N.E.R.C. (1975). Flood Studies Report, vol.1. London: Natural Environment Research Council
6. Otten, A. y Van Montfort, M.A.J., 1978. "The power of two tests on the type of distribution of extremes". J. Hydrology, 37: 195-199.
7. Raynal V., J.A. "La distribución general de valores extremos en la hidrología: genesis, características y propiedades". VIII Congreso Nacional de Hidráulica, Asociación Mexicana de Hidráulica. Toluca, Edo. de Mexico, Octubre 1984. pp. b.1-b.8

8. Tiago de Oliveira, J. (1981) "Statistical choice of univariate extreme-value models. In: G. Taillie et al. (editors), Statistical Distribution in Scientific Work, Vol. 6, D. Reidel, Dordrecht, pp. 653-670
9. Van Montfort, M.A.J. (1970). "On testing that the distribution of extremes is of type I when type II is the alternative". J. Hydrology, 11:421-427
10. Van Montfort, M.A.J. e Ivette Gomes, M. 1985. "Statistical choice of extremal models for complete and censored data" J. Hydrology, 77:77-87

LISTADO DEL PROGRAMA DE COMPUTADOR

```

FILE 5(KIND=DISK,TITLE="GEVRN2",FILETYPE=7)
FILE 6(KIND=REMOTE)
C* GEVRN2:ARCHIVO CON NUMEROS ALEATORIOS
C* N:TAMANO DE LA MUESTRA
C* NSAM:NUMERO DE MUESTRAS
C* X:VARIABLE ALEATORIA (DATA)
C* CONF:NIVEL DE CONFIANZA
C* JF,JW,JG:CONT.VECES STADES IDENTIFICA TIPOS DIST.
C* J1,J2,J3:CONT.VECES VANTES IDENTIFICA TIPOS DIST.
C* XLO:PARAMETRO DE UBICACION
C* XSC:PARAMETRO DE ESCALA
C* XSH:PARAMETRO DE FORMA
C* SBRTN PARGU:CALCULA PARAMETROS DIST.GVE
C* XM:MEDIA DE LAS X'S
C* XS:VARIANZA DE LAS X'S
C* SX:DESVIACION ESTANDAR DE LAS X'S
C* SBRTN STADES:APLICA PRUEBA DE TIAGO DE OLIVEIRA
C* XN0:INVERSA DE LA FUNCION DIST.NORMAL
C* V:VARIABLE DEL ESTADISTICO
C* Z?:VARIABLE ESTANDARIZADA
C* VBAR:MEDIA ARITMETICA DE LAS V'S
C* SDV2:VARIANZA DEL ESTADISTICO CALCULADO(VBAR)
C* TEST:ESTADISTICO DE PRUEBA
C* SDT:VARIANZA DEL ESTADISTICO DE PRUEBA
C* PCL:VALOR DE F(X) DE LA FUNCION NORMAL
C* UCL:VARIABLE CON DISTRIBUCION NORMAL
C* SBRTN HOSTES:APLICA PRUEBA DE HOSKING,WOOD Y WALLIS
C* ZPOS,ZNEG:VALORES TABULADOS DEL ESTADISTICO Z
C* ZT:VALOR TABULADO DE PRUEBA PARA Z CALCULADO
C* SBRTN PWMHWW:ESTIMA PAR.MET.MOM. PROB.PESADA
C* GAMMA:FUNCION GAMMA
C* PU:PARAMETRO DE UBICACION
C* PE:PARAMETRO DE ESCALA
C* PF:PARAMETRO DE FORMA
C* SBRTN MOPROP:CALCULA MOMENTOS DE LAS X'S
C* XM0:MOMENTO DE ORDEN 0
C* XM1:MOMENTO DE ORDEN 1
C* XM2:MOMENTO DE ORDEN 2
C* SBRTN ORDEN:ORDENA LOS VALORES DE X
C* IORD:DE MAYOR A MENOR (1)
C* IORD:DE MENOR A MAYOR (0)
C* SBRTN VANTES:APLICA PRUEBA DE VAN MONTFORT
C* Q:VARIABLE ALEATORIA DATA(COMO X)
C* APOS,ANEG:VALORES TABULADOS DEL ESTADISTICO A
C* DM:MEDIA DE LAS L'S
C* DS:VARIANZA DE LAS L'S
C* ACAL:VALOR CALCULADO DEL ESTADISTICO A
C* ATAB:VALOR TABULADO DE PRUEBA PARA ACAL
C* N=10;CONF=0.95;XLO=10.0;XSC=4.0;XSH=0.10;NSAM=5000
DIMENSION X(100)
JF=0;JW=0;JG=0
JU=0;JD=0;JT=0
J1=0;J2=0;J3=0
DO 3 J=1,NSAM
READ(5,1)(X(I),I=1,N)
1 FORMAT(5E16.10)
DO 2 I=1,N
IF(XSH.NE.0.0)GO TO 15
X(I)=-ALOG(-ALOG(X(I)))*XSC+XLO
GO TO 16
15 X(I)=XLO+XSC*(1.-(-ALOG(X(I)))*XSH)/XSH

```

```

16 CONTINUE
CALL PARGU1(N,X,XSC1,XLO1,2,1,0)
CALL STADES(N,X,XLO1,XSC1,CONF,VBAR,SDV,JF,JW,JG,0)
CALL HOSTES(N,X,0,JU,JD,JI)
CALL VANTES(N,X,0,J1,J2,J3)
3 CONTINUE
WRITE(6,5)NSAM,XSH
5 FORMAT(7,1X,"NUMERO DE MUESTRAS ANALIZADAS=" ,I5,/,1X,
*"PARAMETRO DE FORMA="FS.2)
WRITE(6,4)JG,JF,JW
4 FORMAT(7,1X,"EL CRITERIO DE TIAGO DE OLIVEIRA IDENTIFICO:" ,/,1X,
*"COMO DISTRIBUCION GUMBEL " ,I5," DE LAS MUESTRAS" ,/,1X,
*"COMO DISTRIBUCION FRECHET " ,I5," DE LAS MUESTRAS" ,/,1X,
*"COMO DISTRIBUCION WEIBULL " ,I5," DE LAS MUESTRAS")
WRITE(6,6)JU,JD,JI
6 FORMAT(7,1X,"EL CRITERIO DE HOSKIN, WOOD Y WALLIS IDENTIFICO:" ,/
* ,I5," COMO DISTRIBUCION GUMBEL " ,I5," DE LAS MUESTRAS" ,/,1X,
* ,I5," COMO DISTRIBUCION FRECHET " ,I5," DE LAS MUESTRAS" ,/,1X,
* ,I5," COMO DISTRIBUCION WEIBULL " ,I5," DE LAS MUESTRAS")
WRITE(6,7)J1,J2,J3
7 FORMAT(7,1X,"EL CRITERIO DE VAN MONTFORT IDENTIFICO:" ,/,1X,
*"COMO DISTRIBUCION GUMBEL " ,I5," DE LAS MUESTRAS" ,/,1X,
*"COMO DISTRIBUCION FRECHET " ,I5," DE LAS MUESTRAS" ,/,1X,
*"COMO DISTRIBUCION WEIBULL " ,I5," DE LAS MUESTRAS")
END
SUBROUTINE STADES(N,X,XLO,XSC,CONF,VBAR,SDV,JF,JW,JG,IW)
EXTERNAL XNQ
DIMENSION X(N),DIST(3)
DATA DIST/6HFRECHE,6HWEIBUL,6HGUMBEL/
AN=N
SDT=2.09797;VBAR=0.0;SDV2=0.0
DO 2 I=1,N
Z=(X(I)-XLO)/XSC
V=-Z+(1.-EXP(-Z))*Z**2./2.
VBAR=VBAR+V/AN
2 SDV2=SDV2+V**2./AN
SDV=SDV2-VBAR**2.
PCL=CONF+(1.-CONF)/2.
UCL=XNQ(PCL)
TEST=SQRT(SDT/AN)*UCL
IF(TEST.LT.VBAR)IDIS=1
IF(IDIS.EQ.1)JF=JF+1
IF(-TEST.GE.VBAR)IDIS=2
IF(IDIS.EQ.2)JW=JW+1
IF(VBAR.GT.-TEST.AND.VBAR.LE.TEST)IDIS=3
IF(IDIS.EQ.3)JG=JG+1
IF(IW.EQ.0)RETURN
3 WRITE(6,6)DIST(IDIS),VBAR,TEST,SDV,SDT
6 FORMAT(7,1X,"THE TIAGO DE OLIVEIRA S CRITERION SUGGESTS A " ,A6,
*" DISTRIBUCION" ,/,1X," COMPUTED STATISTIC=" ,F10.6,/,1X," TEST STATI
*STIC=" ,F10.6,/,1X," COMPUTED STATISTIC VARIANCE=" ,F10.6,/,1X,
*"TEST STATISTIC VARIANCE=" ,F10.6,/)
RETURN
END
FUNCTION XNQ(PROB)
IF(PROB.GT.0.5)GO TO 1
PRON=PROB
GO TO 2
1 PRON=1.-PROB
2 V=SQRT(-2.*ALOG(PROB))
XNQ=V-(2.515517+0.802853*V+0.010328*V**2.)/(1.+1.432788*V+0.189269
*V**2.+0.001308*V**3.)
IF(PROB.LT.0.5)XNQ=-XNQ

```

```

RETURN
END
SUBROUTINE PARGU1(N,X,XSC,XLO,IEST,IDIS,IWR)
DIMENSION X(N),EST(2),TRANS(2)
DATA EST/6HBIASED,6HUNBIAS/
DATA TRANS/6HGUMBEL,6HLOGUMB/
VN=N
IF(IDIS.EQ.1)GO TO 20
DO 10 I=1,N
10 X(I)=ALOG(X(I))
20 XM=0.;XS=0.
DO 30 I=1,N
30 XM=XM+X(I)/N
XS=XS+X(I)**2./N
XS=XS-XM**2.
IF(IEST.EQ.2.)XS=XS*VN/(VN-1.)
SX=SQRT(XS)
XSC=0.7806*SX
XLO=XM-0.45*SX
IF(IWR.EQ.0)RETURN
PRINT 40,TRANS(IDIS),EST(IEST),XLO,XSC
40 FORMAT(1H1,///,5X,"PARAMETERS OF THE EXTREME VALUE DISTRIBUTION"
*,/ ,5X,"(TYPE-1, EXTREMAL(" A10,") DISTRIBUTION)",/ ,13X,A9,"MOMENT
*ESTIMATORS)",///,7X,"LOCATION PARAMETER=",F12.5,3X,"(MODE)",7X,
*"SCALE PARAMETER=",F12.5,/)
RETURN
END
SUBROUTINE PWMHWW(N,X,PU,PE,PF,IMP)
EXTERNAL GAMMA
DIMENSION X(N)
IF(IMP.EQ.1)WRITE(6,2)
CALL ORDEN(N,X,IORD)
CALL MOPROP(N,X,XM0,XM1,XM2,IMP)
CONS=(2.*XM1-XM0)/(3.*XM2-XM0)-ALOG(2)/ALOG(3)
PF=7.859*CONS+2.9554*CONS**2.
Z=1.+PF
XZ=GAMMA(Z)
PE=(2.*XM1-XM0)*PF/(XZ*(1.-1./2.**PF))
PU=XM0+PE*(XZ-1.)/PF
IF(IMP.NE.1)RETURN
WRITE(6,4)PU,PE,PF
RETURN
2 FORMAT(///2X"DISTRIBUCION GENERAL DE VALORES EXTREMOS ",//2X"METOD
*O DE MOMENTOS DE PROBABILIDAD PESADA")
4 FORMAT(//2X"PARAMETRO DE UBICACION ="F12.4,//2X"PARAMETRO DE ESCAL
*A ="F12.4,//2X"PARAMETRO DE FORMA ="F12.4)
END
SUBROUTINE ORDEN (N,X,IORD)
DIMENSION X(N)
L=N-1
IF (IORD.GT.0) GO TO 1
DO 2 I=1,L
M=I+1
DO 3 J=I,N
3 IF (X(J).LT.X(M)) M=J
T=X(I)
X(I)=X(M)
2 X(M)=T
RETURN
1 DO 4 I=1,L
M=I+1
DO 5 J=I,N
5 IF (X(J).GT.X(M)) M=J
T=X(I)

```

```

4 X(I)=X(M)
X(M)=I
RETURN
END
SUBROUTINE MOPROP (N,X,XM0,XM1,XM2,IMP)
DIMENSION X(N)
SUM1=0.0
SUM2=0.0
SUM3=0.0
AN=N
DO 1 I=1,N
SUM1=SUM1+X(I)/AN
SUM2=SUM2+X(I)*(I-1)/(AN*(AN-1.))
SUM3=SUM3+X(I)*(I-1)*(I-2)/(AN*(AN-1.)*(AN-2.))
1 CONTINUE
XM0=SUM1
XM1=SUM2
XM2=SUM3
IF (IMP .NE. 1) RETURN
WRITE(6,2)XM0,XM1,XM2
2 FORMAT(//2X"M0="F12.4, /2X"M1="F12.4, /2X"M2="F12.4)
RETURN
END
FUNCTION GAMMA(Z)
IF (Z.GE.57.) GO TO 10
XX=Z
ERROR=1.E-06
XG=1.
GAMMA=XG
IF (XX-2.) 3,9,4
2 IF (XX.LE.2.) GO TO 8
4 XX=XX-1.
XG=XG*XX
GO TO 2
3 IF (XX-1.) 5,9,8
5 IF (XX.GE.ERROR) GO TO 7
XY=FLOAT(INT(XX))-XX
IF (XY.EQ.0.) GO TO 10
IF (ABS(XY).LE.ERROR) GO TO 10
IF ((1.-XY-ERROR).LE.0.) GO TO 10
6 IF (XX.GT.1.) GO TO 8
7 XG=XG/XX
XX=XX+1.
GO TO 6
8 XY=XX-1.
XGX=1.+XY*(-.57710166+XY*(0.98585399+XY*(-.87642182+XY*(
*.8328212+XY*(-.5684729+XY*(.25482049+XY*(-.0514993))))))
GAMMA=XG*XGX
9 RETURN
10 WRITE(6,11)
GAMMA=1.E68
11 FORMAT(//2X"LA FUNCION GAMMA TIENDE A +/- INFINITO")
RETURN
END
SUBROUTINE HOSTES(N1,Q1,IW,JU,JD,JT)
DIMENSION T(80),Q1(100),DIST(3),ZPOS(4),ZNEG(4),TIT(80)
DATA DIST/6HGUMBEL,6HFRECHE,6HWEIBUL/
DATA ZPOS/3.4,4.3,4.6,4.9/
DATA ZNEG/-3.7,-4.6,-4.9,-5.1/
CALL PWHHW(N1,Q1,PU,PE,PF,0)
AN=N1
Z=PF*SQRT(AN/0.5633)
IF (Z.LE.0.) GO TO 12
IF (N1.EQ.10) ZT=ZPOS(1)

```

```

IF(N1.EQ.25)ZT=ZPOS(3)
IF(N1.EQ.50)ZT=ZPOS(3)
IF(N1.EQ.100)ZT=ZPOS(4)
IF(Z.GE.ZT)IDIS=3
IF(Z.LT.ZT)IDIS=1
GO TO 13
12 IF(N1.EQ.10)ZT=ZNEG(1)
IF(N1.EQ.25)ZT=ZNEG(2)
IF(N1.EQ.50)ZT=ZNEG(3)
IF(N1.EQ.100)ZT=ZNEG(4)
IF(Z.LE.ZT)IDIS=2
IF(Z.GT.ZT)IDIS=1
13 IF(IDIS.EQ.1)JU=JU+1
IF(IDIS.EQ.2)JD=JD+1
IF(IDIS.EQ.3)JT=JT+1
IF(IW.EQ.0)RETURN
WRITE(6,5)DIST(IDIS),Z,ZT,AN
5 FORMAT(7,1X,"THE HOSKING-WALLIS-WOOD CRITERION GIVES A",A6,"DISTR
*IBUTION" / 1X,"THE COMPUTED STATISTIC=",F10.6,/,1X,"TABULATED STATIS
*TIC=",F10.6,/,1X,"YRS=",F10.6)
9 CONTINUE
END
SUBROUTINE VANTES(N1,Q,IW,J1,J2,J3)
EXTERNAL XLN
DIMENSION Q(1000),TIT(30),DIST(3)
DIMENSION ANLG(4),APOS(4)
DATA DIST/6HGUMBEL,6HFRECHE,6HWEIBUL/
DATA ANEG/-1.72,-1.83,-1.87,-1.89/
DATA APOS/1.43,1.39,1.37,1.31/
CALL ORDEN(N1,Q,0)
DM=0.;DS=0.;SUM1=0.;SUM2=0.
AN=N1
N=N1
DO 11 I=2,N
K=I-1
UM=XLN(K,AN)
AM=XLN(I,AN)
X=I-.5
DEL=-XLN(X,AN)
DM=DM+DEL/(AN-1)
DS=DS+DEL**2/(AN-1)
ELE=(Q(I)-Q(K))/(AM-UM)
SUM1=SUM1+ELE
11 SUM2=SUM2+DEL*ELE
DS=DS-DM**2
ACAL=(SUM2/SUM1-DM)*SQRT(AN/DS)
IF(ACAL.LE.0.)GO TO 12
IF(N.EQ.10)ATAB=APOS(1)
IF(N.EQ.25)ATAB=APOS(2)
IF(N.EQ.50)ATAB=APOS(3)
IF(N.EQ.100)ATAB=APOS(4)
IF(ACAL.GT.ATAB)IDIS=3
IF(ACAL.LE.ATAB)IDIS=1
GO TO 13
12 IF(N.EQ.10)ATAB=ANEG(1)
IF(N.EQ.25)ATAB=ANEG(2)
IF(N.EQ.50)ATAB=ANEG(3)
IF(N.EQ.100)ATAB=ANEG(4)
IF(ACAL.LT.ATAB)IDIS=2
IF(ACAL.GE.ATAB)IDIS=1
13 IF(IDIS.EQ.1)J1=J1+1
IF(IDIS.EQ.2)J2=J2+1
IF(IDIS.EQ.3)J3=J3+1
IF(IW.EQ.0)RETURN

```

```

6 WRITE(6,6)DIST(IDIS),ACAL,ATAB
6 FORMAT(7,1X,"THE VAN MONTFORT CRITERION SUGGESTS A",A6,"DISTRIBU
*TION" / 1X,"THE COMPUTED STATISTIC IS=",F10.6,/,1X,"THE TABULATED
*STATISTIC IS=",F10.6)
END
FUNCTION XLN(X,AN)
XLN=-ALOG(-ALOG(X/(AN+1)))
RETURN
END

```