UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

APUNTES DE

ANALISIS ESTADISTICO Y PROBABILISTICO DE DATOS HIDROLOGICOS

HIDROLOGIA SUPERFICIAL

ROLANDO SPRINGALL GALINDO

DIVISION DE INGENIERIA CIVIL, TOPOGRAFICA Y GEODESICA
DEPARTAMENTO DE HIDRAULICA

FI/DICTG/86 058

G-601424

I N D I C E

8.	ANALISIS ESTADISTICO Y PROBABILISTICO DE DATOS HIDROLOGICOS	1
8.1	Fundamentos de probabilidad y estadística	1
8.1.1	Distribución de frecuencias y probabilidades	8
8.1.2	Parámetros estadísticos	10
8.1.3	Pruebas F y t	15
8.1.4	Ajuste de distribuciones	20
8.2	Correlación lineal simple	30
8.3	Correlación lineal múltiple	49
8.4	Ordenamiento de datos hidrológicos	63
8.4.1	Selección de un registro	63
8.4.2	Período de retorno	64
8.4.3	Períodos de retorno de los datos	68
8.4.4	Períodos de retornos de los eventos de diseño	70
8.5	Distribuciones de datos hidrológicos	74
8.5.1	Distribución de valores extremos Tipo I	75
8.5.2	Distribuciones ajustadas por mínimos cuadrados	82
8.5.3	Distribución de frecuencias aplicada a dos poblaciones	91

8.6	Extrapolación de las curvas de frecuencias	96
8.7	Análisis de gastos máximos anuales	98
8.8	Análisis de lluvias	100
8.8.1	Distribuciones de alturas de lluvia por una duración constante	101
8.8.2	Curvas de alturas de lluvia-duración-período de retorno	101
8.8.2.1	Criterio propuesto por Chow	103
8.8.2.2	Ajuste por correlación lineal múltiple	105
8.8.3	Ajustes y utilización de las curvas de altura de lluvia-duración-período de retorno	121
8.9	Generación de Información Hidrológica	128
8.9.1	Series Cronológicas	129
8.9.2	Procesos de Generación	133
8.9.2.1	Promedios móviles	133
8.9.2.2	Proceso de suma de armónicas	138
8.9.2.3	Proceso autorregresivo	139

8. ANALISIS ESTADISTICO Y PROBABILISTICO DE DATOS HIDROLOGICOS

En los capítulos anteriores se han indicado los elementos que inter - vienen en el proceso general lluvia-escurrimiento, así como la for - ma de medirlos y cuantificarlos.

Aquí se introducen los conceptos de probabilidad y estadística aplica dos al análisis de datos hidrológicos; primero se plantean los fundamentos de probabilidad y estadística, para después aplicarlos a los registros de datos hidrológicos y proceder al estudio de las técnicas más comunes de probabilidad y estadística utilizadas en el análisis de dichos registros.

8.1 Fundamentos de probabilidad y estadística

A continuación se presenta un breve resumen de los conceptos de es tadística más usuales en hidrología a fin de resaltar fundamentalmente la importancia que tienen estos en la solución de los problemas hidrológicos, sin pretender desarrollar una teoría de los mismos, para lo cual al final del capítulo se proporciona una lista bibliográfica de libros relativos al tema.

Desde el punto de vista estadístico, un registro de datos hidrológicos se conoce como una muestra del comportamiento del fenómeno que se está analizando y midiendo, y al conjunto de todos los posibles registros bajo ciertas condiciones: población o universo del mismo fenómeno. Al proceso de observación o de registro se le llama experimento y las respuestas obtenidas de éste se conocen como resultados. En hidrología, los resultados así definidos tienen una descripción numérica y se les denomina variables aleatorias.

De acuerdo con lo anterior se ve que al analizar cualquier registro de datos hidrológicos, se tiene exclusivamente una muestra de estos y que nunca es factible disponer de la población de ellos, ya que los datos están ligados a fenómenos naturales cuyos resultados son siempre cambiantes. El problema por lo tanto es conocer que tan representativos son los resultados obtenidos de la muestra, del universo de los mismos. Es aquí donde las técnicas estadísticas son de gran utilidad, ya que si la muestra es representativa de la población permite hacer deducciones de esta, en relación con la naturaleza de la población.

El análisis estadístico de datos hidrológicos se puede hacer utilizando alguno de los modelos de distribuciones de probabilidad más usuales o bien, con base en un ajuste de los mismos mediante cierta técnica matemática que es generalmente la de mínimos cuadrados.

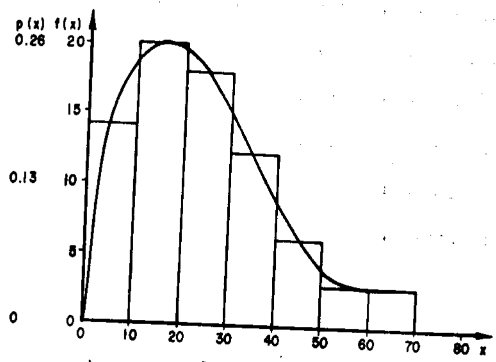
8.1.1 Distribución de frecuencias y probabilidades.

Las variables aleatorias pueden ser discretas o continuas, dependiendo de la forma como se obtengan los resultados; por ejemplo, el número de días de lluvia es una variablea aleatoria discreta, mientras que las alturas de lluvia son variables aleatorias continuas. Sin embargo, para propósitos prácticos una variable discreta puede tratarse arbitrariamente como continua, ajustando una función continua a la variación, o bien una continua como discreta, dividiendo las variables continuas en intervalos y agrupándolos en números discretos.

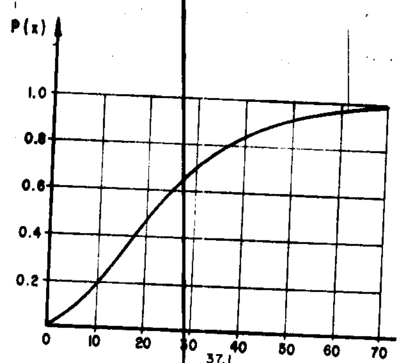
Para conocer la distribucón de frecuencias de una muestra de variables aleatorias discretas es necesario dividir los datos ya obtenidos u observados en intervalos de clase o categorías a fin de valuar la frecuencia de cada intervalo. Si "z" es el número de datos agrupados en cierto intervalo de clase Δx_i, la frecuencia de ese intervalo de clase se define como

$$f_i = \frac{z}{\Delta x_i} \qquad (8.1)$$

Si se grafica la frecuencia f_i correspondiente a cada intervalo de clase Δx_i, se obtiene lo que se conoce como distribución de frecuencias (Fig 8.1); sin embargo, esta distribución sacrifica cierta información contenida en el grupo de datos, ya que se trabaja con intervalos de clase en lugar de los valores específicos de cada una de las variables aleatorias.



a) Distribución de frecuencias y probabilidades



b) Distribución acumulada de probabilidades

Fig. 8.1 Distribuciones Estadísticas

La probabilidad p de que un valor x esté comprendido en un cierto intervalo de clase, se obtiene dividiendo el número de datos "z" contenidos en ese intervalo entre el número total de datos "n" de la muestra. Si el intervalo está acotado entre los valores de a y b, lo anterior se puede expresar como

$$p\ (a \leq x \leq b\ ) = \frac{z_{ab}}{n} \tag{8.2}$$

Sustituyendo la ec 8.1 en la 8.2, se tiene que

$$p\ (a \leq x \leq b\ ) = \frac{1}{n} \sum_{a}^{b} f_i\ \Delta x_i \tag{8.3}$$

lo cual implica que la probabilidad de que x fluctúe entre a y b es el área del histograma comprendido entre esos valores. La probabilidad total de todas las posibles variaciones es igual a uno. Si en lugar de frecuencias se utilizan probabilidades, la distribución obtenida se conoce como distribución de probabilidades; a su vez ésta y la de frecuencias se denominan distribuciones estadísticas.

Para una variable aleatoria continua, la probabilidad de una variación puede considerarse como la probabilidad p (x) de un valor discreto contenido en el intervalo de x a (x + Δx). Como x es un valor continuo, Δx tiende a dx y la probabilidad p (x) llega a ser una función continua denominada densidad de probabilidad. De la ec 8.3, si Δx tiende a dx, se tiene que

$$p\,(a \leq x \leq b) = \frac{1}{n} \int_a^b f\,(x)\,dx = \int_a^b p\,(x)\,dx \tag{8.4}$$

La probabilidad acumulada de una variación (fig 8.1b) es la probabilidad de que la variable aleatoria tenga un valor igual o menor que un cierto valor asignado x. Dicha probabilidad puede designarse como $P\,(X \leq x)$ y expresarse de acuerdo con la ec 8.4 como

$$P\,(X \leq x) = \int_{-\infty}^{x} p\,(x)\,dx \tag{8.5}$$

donde, en este caso, la distribución de probabilidades se considera ilimitada. Si el límite superior de la integral $x = \infty$, entonces $P\,(X \leq x) = 1$, lo que constituye el evento seguro. Si la distribución de probabilidades es limitada, o sea que la densidad de probabilidades p (x) se define para un cierto intervalo $(a \leq x \leq b)$, la ecuación anterior también es válida, considerando que $p\,(x) = 0$ para todos los valores de x fuera del intervalo especificado.

En la Tabla 8.1 se muestran las distribuciones de probabilidades más usuales, así como sus parámetros más importantes.

Ejemplo 8.1 Determinar la distribución de frecuencias, de probabilidades y su acumulada, del registro de lluvias mostrado en la Tabla 8.2. Obtener la altura de lluvia para una probabilidad acumulada de 0.8.

Para conocer la distribución de frecuencias de las alturas de lluvia

TABLA 8.1

DISTRIBUCIONES DE PROBABILIDADES MAS USADAS EN HIDROLOGIA

Nombre de la distribución	EXPRESION MATEMATICA	Media	Variancia	Nomenclatura
POISSON	$p(x) = \frac{m^x e^{-m}}{x!}$ $m = np$	m	m^2	a – parámetro b – parámetro c – parámetro n – número total de datos N – frecuencia total p – probabilidad de ocurrencia de un evento p(x) – densidad de probabilidad p(z) – densidad de probabilidad x – variable y – variable z – variable normalizada μ – media de las x μ_2 – segundo momento μ_3 – tercer momento μ_4 – cuarto momento μ_y – media de las y σ – desviación estandar de las x σ_y – desviación estandar de las y
NORMAL	$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$; $z = \frac{x-\mu}{\sigma}$	0	1	
GAMMA	$p(x) = \frac{x^a e^{-x/b}}{b^{a+1}\Gamma(a+1)}$ $b > 0,\ x > -1$, para $x = 0$ $p(x) = 0$ para $x \leq 1$ $\Gamma(a+1) = a!$	$b(a+1)$	$b^2(a+1)$	
PEARSON Tipo III	origen en el modo $2\beta_2 = 3\beta_1 + 6$ $p(x) = p_o \left(1 + \frac{x}{a}\right)^c e^{-cx/2}$ $c = \frac{4}{\beta_1} - 1$; $a = \frac{c}{2}\frac{\mu_3}{\mu_2}$ $p_o = \frac{N}{a}\frac{c^{c+1}}{e^c\,\Gamma(c+1)}$; $_1 = \frac{\mu_3^2}{\mu_2^3}$; $_2 = \frac{\mu_4}{\mu_2^2}$	modo – $\mu_3/2\mu_2$	$\sqrt{\mu_2}$	
EXTREMA Tipo I	$p(x) = \frac{1}{c} e^{-(a+x)/c - e^{-(a+x)/c}}$	$0.5772c - a$	$\frac{\pi c}{\sqrt{6}}$	
LOGNORMAL	$p(x) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-(y-\mu_y)^2/2\sigma_y}$ $y = \ln x$	$e^{\mu_y + \sigma_y^2/2}$	$\mu(e^{\sigma_y^2} - 1)^{1/2}$	

(Tabla 8.2) se procedió a seleccionar un intervalo de clasificación, que de acuerdo con el tamaño de la muestra se consideró conveniente de 10 mm. Para efectos de cálculo, a todos los valores de la altura de lluvia entre 0.1 y 10 se les asignó un valor de 5, entre 10.1 y 20 de 15, etc (Tabla 8.3, cols 1 y 2). Seleccionados los intervalos se procedió a contar cuantos valores caen dentro de cada uno de estos, para valuar su frecuencia de acuerdo con la ec 8.1 (Tabla 8.3, col 3).

En la fig 8.1a se muestra la distribución de frecuencias o histograma de las alturas de lluvia, la cual se obtuvo graficando la pareja de valores obtenidos de la tabla 8.3, cols 2 y 3. (histograma).

Aplicando la ec 8.2 a los datos de la tabla 8.3, col 3, se tiene en la col 4, el valor de la probabilidad, el que al vertir en gráfica, respecto a la col 1, se obtiene la distribución de probabilidad de las lluvias analizadas (fig 8.1a). Finalmente, sumando los valores de la probabilidad obtenida en la col 4, se deduce la probabilidad acumulada de los valores de la lluvia analizada, col 5, cuya gráfica se indica en la fig 8.1b.

De la fig 8.1b se deduce que si p = 0.8 se obtiene una altura de lluvia de 37.1 mm.

Esto implica que la probabilidad de que se presente un valor menor o igual a 37.1 es de 80 por ciento, así:

$$P(x \leq 37.1) = 0.80$$

TABLA 8.2 Alturas de lluvias con duración de 6 horas

No. Orden	hp (mm)	No. Orden	hp (mm)	No. Orden	hp (mm)	No. Orden	hp (mm)
1	10.9	20	8.7	39	21.3	58	5.7
2	14.6	21	2.9	40	17.2	59	15.5
3	20.6	22	8.9	41	36.1	60	18.5
4	11.9	23	32.3	42	49.5	61	30.7
5	64.0	24	56.6	43	20.0	62	10.5
6	20.9	25	3.6	44	4.6	63	3.5
7	32.0	26	11.0	45	27.3	64	13.9
8	28.8	27	52.5	46	49.7	65	4.3
9	2.6	28	46.3	47	34.6	66	22.6
10	35.8	29	5.2	48	49.0	67	45.0
11	17.7	30	42.5	49	25.3	68	10.0
12	38.4	31	17.5	50	27.2	69	9.6
13	16.3	32	64.5	51	38.8	70	22.7
14	6.8	33	28.5	52	52.4	71	23.0
15	17.0	34	19.6	53	11.5	72	26.2
16	6.8	35	30.9	54	15.8	73	31.7
17	25.3	36	27.5	55	70.00	74	29.9
18	36.3	37	10.3	56	24.0	75	25.9
19	16.5	38	21.1	57	16.5	76	34.5

TABLA 8.3 Análisis de Frecuencias y probabilidades.

1	2	3	4	5
ırcas de Clase	Intervalo de clase Δxi	Frecuencia $\Delta xifi$	Probabilidad pi	Probabilidad acumulada Pi
5	0.1 - 10	14	0.18	0.18
15	10.1 - 20	20	0.26	0.44
25	20.1 - 30	18	0.24	0.68
35	30.1 - 40	12	0.16	0.84
45	40.1 - 50	6	0.08	0.92
55	50.1 - 60	3	0.04	0.96
65	60.1 - 70	3	0.04	1.00

8.1.2 Parámetros estadísticos

Las características fundamentales de una distribución estadística se pueden conocer a partir de sus parámetros. A continuación se indican los más importantes en hidrología.

La media aritmética o media de una muestra de un conjunto de valores se define como

$$\overline{x} = \frac{\sum_{i=1}^{n} xi}{n} \qquad (8.6)$$

donde xi son los valores de la muestra y n el número total de estos. Para referirse a la media del universo o población se utiliza la letra μ . La media permite conocer la tendencia central de una distribución estadística.

En el caso de una distribución de frecuencias, como se pierde información al trabajar con intervalos de clase en lugar de los datos originales, la media de dicha distribución se valúa como

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i \, fi \, \Delta xi}{n} \qquad (8.7)$$

donde x_i es el valor representativo de un intervalo de clase Δxi, fi su frecuencia correspondiente, k el número total de intervalos de clasificación y n el número total de datos.

Para conocer que tan dispersos están los valores de una muestra respecto de su media, se utiliza lo que se conoce como la desviación estandar de la muestra, la cual se expresa como

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}} \qquad (8.8)$$

donde x_i son los valores de las muestras, $\overline{x}$ su media (ec 8.6) y n el número total de valores de la muestra.

Análogamente, la desviación estandar de una distribución de frecuencias se valúa como

$$S = \sqrt{\frac{\sum_{i=1}^{K} (x_i - \overline{x})^2 \; f_i \quad \Delta x_i}{n-1}} \qquad (8.9)$$

donde las variables tienen el mismo significado que en la ec 8.7. En caso de que la desviación estandar se refiera al universo, esta se indica con la letra σ . Al cuadrado de la desviación estandar se le llama variancia; para una muestra se designa con S^2 y para el universo con σ^2.

La relación entre la desviación estandar y la media se conoce como coeficiente de variación C_v.

Ejemplo 8.2 Calcular los parámetros estadísticos principales de la distribución obtenida en el ejemplo 8.1, trabajando con los valores específicos y con los intervalos de clase.

a) Análisis usando los valores específicos de las alturas de lluvia.

Para facilitar el cálculo se elaboró la Tabla 8.4

De la Tabla 8.4, col 2 y aplicando la ec 8.6, se tiene que la media de los valores específicos de la altura de lluvia es

$$\bar{x} = \frac{1888.10}{76} = 24.84$$

De la col 4, con la ec 8.8, se deduce una variancia de

$$S^2 = \frac{1}{76-1} (19,142.24) = 255.23$$

siendo su desviación estandar y coeficiente de variación de

$$S = \sqrt{S^2} = 15.98$$

$$C_V = \frac{S}{\bar{x}} = 0.64$$

b) Análisis considerando intervalos de clase.

Para la obtención de las características de la distribución de frecuencias de las lluvias se elaboró la Tabla 8.5

De la Tabla 8.5, col 3, se tiene, aplicando la ec 8.7, que la media de la distribución de frecuencia de lluvias es de

$$\bar{x} = \frac{1870}{76} = 24.605 \doteq 24.61$$

De la col 6, y la ec 8.9, se deduce que la variancia de la distribución de frecuencias de lluvias resulta de

$$S^2 = \frac{1}{76-1} (18,688.16) = 249.18$$

Tabla 8.4 Ordenamiento para el cálculo de las características de la muestra (ejemplo 8.2)

NO. ORDEN	Xi	Xi - $\overline{X}$	$(Xi - \overline{X})^2$
1	10.90	-13.94	194.32
2	14.60	-10.24	104.85
3	20.60	-4.24	17.97
4	11.90	-12.94	167.44
5	64.00	39.16	1533.50
6	20.90	-3.94	15.50
7	32.00	7.16	51.26
8	28.80	3.96	15.68
9	2.60	-22.24	494.61
10	35.80	10.96	120.10
11	17.70	-7.14	50.97
12	38.40	13.56	183.87
13	16.30	-8.54	72.93
14	6.80	-18.04	325.44
15	17.00	-7.84	61.46
16	6.80	-18.04	325.44
17	25.30	0.46	0.21
18	36.30	11.46	131.33
19	16.50	-8.34	69.55
20	8.70	-16.14	260.49
21	2.90	-21.94	481.36
22	8.90	-15.94	254.08
23	32.30	7.46	55.65
24	56.60	31.76	1008.69
25	3.60	-21.24	451.13
26	11.00	-13.84	191.54
27	52.50	27.66	765.07
28	46.30	21.46	460.53
29	5.20	-19.64	385.72
30	42.50	17.66	311.87
31	17.50	-7.34	53.87
32	64.50	39.66	1572.91
33	28.50	3.66	13.39
34	19.60	-5.24	27.45
35	30.90	6.06	36.70
36	27.50	2.66	7.07
37	10.30	-14.54	211.41
38	21.10	-3.74	13.98
39	21.30	-3.54	12.53
40	17.20	-7.64	58.36
41	36.10	11.26	126.78
42	49.50	24.66	608.11
43	20.00	-4.84	23.42
44	4.60	-20.24	409.65
45	27.30	2.46	6.05
46	49.70	24.86	618.01
47	34.60	9.76	95.25
48	49.00	24.16	583.70
49	25.30	0.46	0.21
50	27.20	2.36	5.56
51	38.80	13.96	194.88
52	52.40	27.56	759.55
53	11.50	-13.34	177.95
54	15.80	-9.04	81.72
55	70.00	45.16	2039.42
56	24.00	-0.84	0.70
57	16.50	-8.34	69.55
58	5.70	-19.14	366.33
59	15.50	-9.34	87.23
60	18.50	-6.34	40.19
61	30.70	5.86	34.33
62	10.50	-14.34	205.63
63	3.50	-21.34	455.39
64	13.90	-10.94	119.68
65	4.30	-20.54	421.89
66	22.60	-2.24	5.01
67	45.00	20.16	406.42
68	10.00	-14.84	220.22
69	9.60	-15.24	232.25
70	22.70	-2.14	4.57
71	23.00	-1.84	3.38
72	26.20	1.36	1.84
73	31.70	6.86	47.05
74	29.90	5.06	25.60
75	25.90	1.06	1.12
76	34.50	9.66	93.31
Suma	1,888.10		19,142.24

TABLA 8.5 PROCESAMIENTO DE CALCULO (ejemplo 8.2)

x_i	$fi\Delta x_i$	$x_i fi \Delta x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 fi \Delta x_i$
1	2	3	4	5	6
5	14	70	-19.60	384.16	5,378.24
15	20	300	- 9.60	92.16	1,843.20
25	18	450	0.40	0.16	2.88
35	12	420	10.40	108.16	1,297.92
45	6	270	20.40	416.16	2,496.96
55	3	165	30.40	924.16	2,772.48
65	3	195	40.40	1,632.16	4,896.48
SUMA	76	1,870		3,557.12	18,688.16

y la desviación estándar $S = \sqrt{S^2} = 15.78$ y $C_V = \frac{15.78}{24.61} = 0.64$

c) Comparación de los resultados

La comparación de los valores obtenidos de los principales parámetros estadísticos del ejemplo 8.1, trabajando con los valores específicos de las alturas de lluvia y con sus intervalos de clase respectivos se muestra en la tabla siguiente:

PARAMETROS \ RESULTADOS CON	(a) valores específicos	(b) intervalos de clase
Media ($\bar{X}$)	24.84	24.61
Desv.Estándar(S)	15.98	15.78
Variancia (S^2)	255.23	249.18
Coef. Variación (C_V)	0.64	0.64

De la comparación de estos valores se deduce que el empleo de cualquiera de las técnicas presentadas conduce a resultados satisfactorios que aparentemente, no modifican de forma significativa la presición de los mismos. Esto se debió a que había una buena concordancia en la distribución de los valores utilizados en el ejemplo. Conforme se incrementa el intervalo de clase y existe dispersión en los datos se incrementa la diferencia de las características de la distribución al aplicar los criterios antes mencionados.

8.1.3 Pruebas F y t

La forma de la variación de una variable se puede utilizar para conocer los cambios físicos ocurridos en la cercanía de la misma. Para determinar si dos grupos de datos n_1 y n_2, son de una misma población o de diferentes poblaciones con distribución normal, se utiliza la relación de sus variancias S_1^2 y S_2^2, denominada también prueba F, donde

$$F = \frac{S_1^2}{S_2^2} \qquad (8.10)$$

así F es el valor de una variable aleatoria que tiene distribución F con parámetros $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$. Los valores tabulados de la distribución F muestran los valores de esta, que pueden ser excedidos con una probabilidad de 0.05 ó 0.01, valores que constituyen el nivel de significancia, en términos de los grados de libertad

ν_1 y ν_2.

Si el valor de F obtenido de la ec 8.10 excede del valor tabular para el nivel de significancia seleccionada, la diferencia entre las variaciones de los dos grupos o muestras analizadas es significativo en un sentido estadístico, pudiendose concluir la igualdad o no de las poblaciones a que pertenecen. La estadística de la media y de la desviación estándar de una muestra, se pueden utilizar también para conocer la longitud de un registro hidrológico, de tal forma que la media de los datos esté comprendida entre ciertos límites seleccionados de la media de la población. Para esto, se utiliza la distribución t, cuya variable aleatoria t con n-1 grados de libertad está dada por la ec.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{8.11}$$

con intervalo de confianza

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}}$$

En las tablas de la distribución de t, se muestra el valor de t para diversos grados de libertad $\nu = n-1$ y niveles de significancia α. De la ec 8.11 se puede obtener el valor de n procediendo por tanteos, ya que t es función de n. Esta ecuación permite hacer comparación de medias, similares a las expuestas para las variancias, así como conocer la variación de la media de la población a partir de la media de la muestra.

Ejemplo 8.3 Determinar si los registros de lluvia de la estación climatológica de Tepames, Col. son de la misma población, ya que la estación sufrió un cambio en sus registros en 1954, según se vió en el ejemplo 3.5.

Como en 1954 los registros sufrieron un cambio, se procede a determinar si los correspondientes al ciclo 1948 a 1953 són de la misma población que los tomados de 1954 a 1959. Para esto se calcula las características de cada una de las muestras y posteriormente se aplica la prueba F.

De las columnas 2, y aplicando la ec 8.6, la media ($\overline{P}$) de las precipitaciones anuales resulta

Muestra 1 (1948 - 1953)	Muestra 2 (1954 - 1959)
$\overline{P}_1 = \frac{5693.4}{6} = 948.90$	$\overline{P}_2 = \frac{3707.7}{6} = 617.95$

De las columnas 4, y aplicando la ec 8.8, la variancia (S^2) de las precipitaciones anuales para cada una de las muestras resulta

Muestra 1 (1948 - 1953)	Muestra 2 (1954 - 1959)
$S_1^2 = \frac{1}{6-1} (334{,}425.42) = 66{,}885.08$	$S_2^2 = \frac{1}{6-1} (433{,}774.64) = 86{,}754.92$

Con estos valores, aplicando la prueba F (ec 8.10) se obtiene

$$F = \frac{66885.08}{86754.92} = 0.771 = F \text{ cal (F calculada)}$$

MUESTRA 1			
Año	Pa (mm)	Pa - $\bar{P}_1$	$(Pa - \bar{P}_1)^2$
1	2	3	4
1948	1167.90	219.00	47961.00
1949	754.60	-194.30	37752.49
1950	759.70	-189.20	35796.64
1951	1088.20	139.30	19404.49
1952	1272.30	323.40	104587.56
1953	650.70	-298.20	88923.24
n 6	5693.40		334425.42

MUESTRA 2			
Año	Pa (mm)	Pa - $\bar{P}_2$	$(Pa - \bar{P}_2)^2$
1	2	3	4
1954	359.80	-258.15	66641.4
1955	1151.00	533.05	284142.30
1956	714.90	96.95	9399.30
1957	508.90	-109.05	11891.90
1958	603.10	-14.85	220.52
1959	370.00	-247.95	61479.20
n 6	3707.70		433774.64

De los valores tabulados para la distribución F, con grados de libertad $\nu_1 = \nu_2 = 6-1 = 5$, a fin de que la hipótesis de igualdad de varıancias se verifique, es decir que ambas muestras pertenezcan a una misma población, el valor de F que no puede ser excedido con un nivel de significancia $\alpha = 0.01$ es $F_{0.01} = 11.00$, por lo que se concluye que siendo F tab $>$ F cal ambas muestras son de la misma población o bien el valor F con un nivel de significancia de $\alpha = 0.05$ y $\nu_1 = \nu_2 = 5$

$$F_{0.05} = 5.05 = F \text{ tab} \quad (F \text{ tabulada})$$

por lo tanto

obteniéndose la misma conclusión.

Ejemplo 8.4 Si en la Estación Climatológica de Calnali, Hgo., de un registro de 43 años, se tiene una media de las alturas de lluvia máximas anuales diarias de 111.5 mm con una desviación estandar de 25.1 mm. ¿Cuántos años de registro se requieren, para que con una probabilidad de 95 por ciento, la media de la muestra esté dentro del 5 por ciento de la media verdadera?

Lo anterior implica que μ varíe entre 0.95 $\bar{x}$ y 1.05 $\bar{x}$. De la ec. 8.11 se tiene que

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

que define las fluctuaciones de la media del universo en base a la media de la muestra.

Entonces,

$$\frac{t_{\alpha/2}\, s}{\sqrt{n}\ \bar{x}} = 0.05$$

de donde

$$n = \frac{(t_{\alpha/2})^2 \ (s/\bar{x})^2}{(0.05)^2}$$

Como el valor de t es función de n, esta ecuación se tiene que resolver por tanteos; afortunadamente, la fluctuación de t para valores de n mayores de 30 años **es pequeña**. Así, para 95 por ciento de probabilidades, el nivel de significancia es $\alpha = 1-0.95 = 0.05$ y $\alpha/2 = 0.025$ y $t_{\alpha/2} = t_{0.975}$; de una tabla de esta distribución se obtiene que para $\nu = 30-1$ grados de libertad, $t_{0.975} = 2.04$ y para un número infinito de grados de libertad, $t_{0.975} = 1.96$. Considerando $t_{\alpha/2} = 1.96$ y los valores dato de la muestra, se tiene, sustituyendo en la ecuación anterior, que

$$n = \frac{(1.96)^2 \quad (25.1/111.5)^2}{(0.05)^2} = 77.\underline{87} = 78 \text{ años}$$

lo que implica que se requieren 78 años de registro para que la media de la muestra de los datos esté dentro del 5 por ciento de la media verdadera o de la población.

8.1.4 Ajuste de distribuciones

Como se indicó al principio del inciso, si se dispone de una muestra hidrológica de datos de tamaño n y se desean conocer sus propiedades estadísticas, la manera de lograrlo es aceptando a priori que dicha muestra tiene una cierta distribución de probabilidades conocidas, y de ahí inferirlas. Si se hace esto, antes de proceder a utilizar y analizar dicha distribución se requiere conocer que tan cierto es que la distribución elegida se pueda utilizar como representativa del conjunto de datos o muestra disponible.

En estadística existen criterios y técnicas para probar lo anterior, dependiendo del tamaño de la muestra, e inferir cual es la distribución más adecuada para una muestra de datos:

a) Prueba χ^2

Conviene utilizarla cuando el número de observaciones es grande.

Si se considera que f_i y e_i son respectivamente para los k intervalos las frecuencias observadas del fenómeno y las esperadas teóricamente de acuerdo con la ley de distribución de probabilidad escogida como representativa del fenómeno; entonces el valor:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \tag{8.12}$$

tiene distribución χ^2 con ($\nu = k - 1 - m$), grados de libertad; en las tablas de distribución χ^2 , se muestran los valores de esta variable para diversos grados de libertad y niveles de significancia.

Si al utilizar la ec 8.12, el valor calculado es mayor que el tabular, se deduce que la distribución elegida como modelo de las frecuencias observadas de los resultados que se están analizando, no es la correcta; si es menor, se acepta.

Al utilizar esta prueba se debe tener cuidado de que en cada intervalo de clase se tengan por lo menos cinco observaciones.

b) Prueba de Kolmogorov - Smirnov (K & S)

Esta prueba permite hacer comparaciones entre dos distribuciones acumuladas y conviene utilizarla cuando el tamaño de la muestra es pequeña. Para aplicar esta prueba, se grafican tanto los valores de la muestra en orden creciente asignándole a cada valor la misma probabilidad, así como la distribución acumulada de la función con la cual se desea ver, si es o no representativa de la muestra. Una vez graficadas las dos distribuciones, se valúa la máxima ordenada D existente entre ellas.

Para determinar si la diferencia D está dentro del intervalo aceptable y la distribución elegida es la correcta, se calcula de los valores tabulados el valor máximo correspondiente a la prueba de Kolmogorov - Smirnov, de acuerdo con el tamaño de la muestra y el nivel de significancia.

Si el valor determinado de la diferencia en ordenada o frecuencia máxima entre las dos distribuciones, excede de la diferencia tabulada, se deduce que no es correcta la distribución elegida como representativa de los datos y será necesario probar el ajuste para otras distribuciones

Ejemplo 8.5 Aplicación de la prueba χ^2

En la tabla 8.6, se presentan los registros de gastos máximos aforados del río de la Laja en la estación hidrométrica Pericos, Gto., sobre la -

cuenca Lerma - Santiago. Se desea conocer si los valores tienen una distribución estadística normal utilizando la prueba χ^2.

Para aplicar la prueba χ^2 al registro de gastos máximos, se requiere ajustar éstos considerando intervalos de clase. Dado el rango de valores de los gastos máximos se consideraron 4 marcas de clase. La variación de las marcas de clase así como el estudio correspondiente se muestra en la tabla 8.7.1

De la ec (8.6) se deduce para los valores mostrados en la tabla (8.6) que la media de los gastos es:

$$\overline{X} = \frac{5186.8}{34} = 152.6$$

Y de la ecuación (8.7), para la distribución de frecuencias (Tabla 8.7.1) se deduce que:

$$\bar{x} = \frac{5300}{34} = 155.9$$

Como no existe una diferencia significativa en los valores obtenidos, se concluye que los intervalos de clase elegida son los apropiados.

En la Tabla 8.7.2 se muestra el cálculo de las frecuencias esperadas de acuerdo a una distribución normal. Así, en las cols 1 y 2 se indican respectivamente los intervalos de clase elegidos para la distribución de los gastos y los límites de clase asociados a cada intervalo.

Tabla 8.6 Gastos máximos registrados en la Estación Pericos, Gto.

Orden	Año	Q máx
1	1929	16.40
2	1930	188.00
3	31	228.00
4	32	114.00
5	33	170.00
6	34	134.00
7	35	312.00
8	36	222.00
9	37	230.00
10	38	219.00
11	39	193.00
12	1940	109.00
13	41	174.00
14	42	108.00
15	43	231.00
16	44	335.00
17	45	141.00

Orden	Año	Q máx
18	46	30.90
19	47	105.00
20	48	84.10
21	49	49.70
22	1950	64.20
23	51	125.00
24	52	93.40
25	53	182.00
26	54	92.90
27	55	381.00
28	56	129.00
29	57	33.80
30	58	224.00
31	59	92.60
32	1960	131.00
33	61	81.80
34	1962	162.00
SUMA		5186.80

Tabla 8.7.1 Análisis de Frecuencias.

Marca de Clase (Xi)	Intervalo de Clase (ΔXi)	Frecuencia (fi ΔXi)	Xifi ΔXi
50	0-100	10	500
150	101-200	15	2250
250	201-300	6	1500
350	301-400	3	1050
SUMA		34	5300

Para obtener los valores estandarizados para los límites de clase correspondientes, se aplica la ecuación:

$$z = \frac{x - \bar{x}}{s}$$

donde x es el límite de clase, $\bar{x}$ es la media de los gastos máximos y s la desviación estandar de los gastos máximos deducida de la distribución de frecuencias.

De la ec (8.9) para la información mostrada en la Tabla 8.7.1, se obtiene una desviación estandar de la muestra de 91.92, con lo cual considerando $\bar{x}$ = 155.90, la ecuación anterior se transforma a

$$z = \frac{x - 155.90}{91.92}$$

cuyos valores para los límites de clase se indican en la Tabla 8.7.2, col 3. En las cols 4 y 5 se tiene respectivamente el área bajo la curva de la distribución normal correspondientes a los valores de z encontrados y el área para cada intervalo de clase. Esta última se deduce sumando o restando, los dos valores sucesivos de la col 4, dependiendo del signo de la variable z; con igual signo se restan y con signo diferente se suma. Estos valores representan la frecuencia relativa de ocurrencia del evento.

Multiplicando la frecuencia relativa (col 5) de cada evento por el número total de ellos (n = 34) se obtiene la frecuencia esperada para los valores ajustados a la distribución normal (col 6). Finalmente en la col 7 se tiene la frecuencia observada, la cual se dedujo en la Tabla 8.7.1. En la Tabla 8.7.3 se ordenan las frecuencias antes indicadas para la aplicación de la ec 8.12.

Así, se deduce que

$$\chi^2_{cal} = 2.5162 \doteq 2.52$$

De las Tablas χ^2 se deduce para $\nu = k-1-m=4-1-2 =$ grados de libertad y un nivel de significancia $\alpha = 5\%$ que

$$\chi^2_{0.95} = 3.84; \text{ y para } \alpha = 1\%, \ \chi^2_{0.99} = 6.63.$$

En ambos casos $\chi^2_{cal} < \chi^2_{tab}$ por lo que se puede concluir que para los niveles de significancia probados, los gastos máximos anuales registrados de la Estación Pericos, Gto. sobre el Río de la Laja tienen una distribución normal.

<u>Ejemplo 8.6</u> Aplicación de la prueba de Kolmogorov & Smirnov.

Comprobar la validez del ajuste a la distribución normal de los gastos máximos en el Río de La Laja, Gto., analizada en el ejemplo anterior. Para valuar la diferencia máxima en ordenada "D" que existe entre las distribuciones observada y teórica, en la Tabla 8.8 se proporciona el ordenamiento de los datos Tabla 8.6 y su frecuencia acumulada. En la

Tabla 8.7.2 Cálculo de la Frecuencia Esperada Considerando Distribución Normal.

1	2	3	4	5	6	7
AXi	Límites de clase X	Z=F (X)	Area baja la Curva Normal de o a z	Area para Cada intervalo de clase	Frecuencia Esperada	Frecuencia Observada
	0	-1.6960	0.4550			
0-100	100.50	-0.6026	0.2341	0.2209	7.5-8	10
101-200	200.50	0.4852	0.1862	0.4203	14.2-14	15
201-300	300.50	1.5731	0.4422	0.2560	8.7-9	6
301-400	400.50	2.6610	0.4961	0.0539	1.8-2	3

Tabla 8.7.3 Ordenación de cálculo para la aplicación de la Prueba "χ^2".

	f_i	e_i	$(f_i - e_i)$	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
1	10	7.5	2.5	6.25	0.8333
2	15	14.2	0.8	0.64	0.0450
3	6	8.7	-2.7	7.29	0.8379
4	3	1.8	1.2	1.44	0.8000
SUMA					2.5162

fig 8.2 se muestra el histograma acumulado de los valores de la muestra, ordenados de menor a mayor (Tabla 8.8 col. 3) asignándole a cada uno la misma probabilidad o frecuencia 1/n (Tabla 8.8, col. 1) y la distribución acumulada de frecuencia de los valores esperados correspondientes a la distribución normal teórica (Tabla 8.7.2 cols 1 y 6) con - $P(X \leq x)$. De la fig 8.2 se deduce la diferencia en ordenada máxima entre (a) y (b): D = 0.105. De la tabla de distribución K & S, con n=34 y α = 0.01 y 0.05, D^* = 0.200. Como $D < D^*$ se acepta la distribución normal como representativa de los gastos máximos anuales.

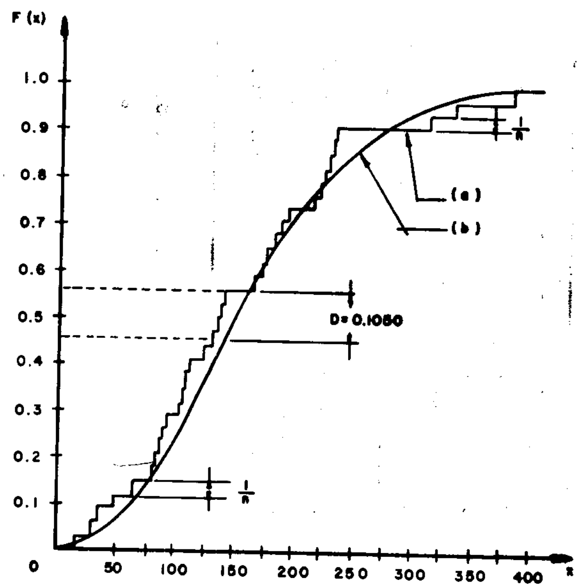


Fig. 8.2 Distribuciones Acumuladas

Tabla 8.8 Ordenación de los datos y frecuencias acumuladas.

Orden (i)	Q máx (Xi) m^3/seg	Xi de menor a mayor	F (Xi) (i/n)
1	2	3	4
1	16.40	16.40	0.03
2	188.00	30.90	0.06
3	228.00	33.80	0.09
4	114.00	49.70	0.12
5	170.00	64.20	0.15
6	134.00	81.80	0.18
7	312.00	84.10	0.21
8	222.00	92.60	0.24
9	230.00	92.90	0.26
10	219.00	93.40	0.29
11	193.00	105.00	0.32
12	109.00	108.00	0.35
13	174.00	109.00	0.38
14	108.00	114.00	0.41
15	231.00	125.00	0.44
16	335.00	129.00	0.47
17	141.00	131.00	0.50
18	30.90	134.00	0.53
19	105.00	141.00	0.56
20	84.10	162.00	0.59
21	49.70	170.00	0.62
22	64.20	174.00	0.65
23	125.00	182.00	0.68
24	93.40	188.00	0.71
25	182.00	193.00	0.74
26	92.90	219.00	0.76
27	381.00	222.00	0.79
28	129.00	224.00	0.82
29	33.80	228,00	0.85
30	224.00	230.00	0.88
31	92.60	231.00	0.91
32	131.00	312.00	0.94
33	81.80	335.00	0.97
34	162.00	381.00	1.00

8.2 Correlación lineal simple

El análisis de correlación se utiliza para conocer cómo una variable independiente (x) afecta a una variable dependiente (y). Si existe solo una variable independiente involucrada, al proceso de análisis se le conoce como de correlación simple; si hay más de una variable independiente, se denomina correlación múltiple.

Es conveniente antes de proceder a un análisis de correlación simple de una serie de parejas de datos, graficar estos, con el fín de conocer la tendencia de la naturaleza en la relación de los datos. Si su forma tiende a una línea recta, la relación se dice lineal, si es curva, la relación se denomina curvilínea. Esta última es factible cambiarla a lineal realizando transformaciones de los ejes coordenados.

Una vez graficadas las parejas de valores de datos y conocida su tendencia, se correlacionan para conocer cual es la relación que mejor se ajusta a dicha tendencia. Por tanto, si la tendencia es una línea recta, a la cual se denomina recta de regresión, para calçular su ecuación, se puede utilizar el método de mínimos cuadrados.

En la fig 8.3 se muestra una serie de parejas de datos (x_i, y_i) en cuya correlación simple, la ecuación de la recta de regresión se puede escribir como

$$y_i' = a + b\ x_i \tag{8.13}$$

la cual plantea el problema de calcular los valores de los parámetros a y b, tales que proporcionen el mejor ajuste de los datos. Para esto, como ya se indicó se puede utilizar el método de los mínimos cuadrados, el cual se basa en que la suma de los errores al cuadrado sea mínimo.

El error (e) para cada punto muestreado se obtiene como

$$e_i = y_i - (a + b\ x_i) \tag{8.14}$$

donde y_i es el valor dato, y $(a + b\ x_i)$ es el valor inferido u obtenido de la ecuación de la recta de regresión.

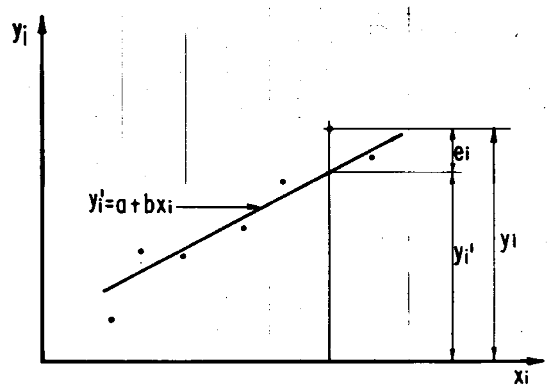


Fig. 8.3 Correlación simple

Haciendo que la sumas de los errores (ei) al cuadrado de cada punto dato ec (8.14) sea mínimo, se obtienen las ecuaciones simultáneas

$$\sum_{i=1}^{n} y_i = an + b \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2$$

donde

n número de pareja de dato.

Se tienen entonces dos ecuaciones con dos incógnitas a y b que son los parámetros buscados. De otra forma, demuestra que pueden calcularse según se relacionen.

$$b = \frac{Sxy}{Sxx} \tag{8.15}$$

y

$$a = \bar{y} - b.\bar{x} \tag{8.16}$$

donde

$$Sxx = n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \tag{8.17}$$

$$Sxy = n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) \quad (8.18)$$

y $\bar{x}$ y $\bar{y}$ son la media de los valores x_i y y_i respectivamente - (ec. 8.6).

La ecuación de la recta de regresión (ec 8.13) así obtenida es para cada x_i, la media de la variación de la variable dependiente y_i. Conforme la pareja de los valores x_i, y_i tiendan a agruparse sobre una línea recta la variancia del error e_i tenderá a cero. La variancia del error se puede escribir como

$$Se^2 = \frac{Syy}{n(n-2)}\left[1 - \frac{(Sxy)^2}{SxxSyy}\right] = Sy^2\left[1 - r_{xy}^2\right] \quad (8.19)$$

siendo

$$r_{xy} = \frac{Sxy}{(Sxx\,Syy)^{1/2}}$$

donde S_{yy} es una expresión similar a la ec (8.17) sustituyendo a las x_i por y_i, S_y^2 la variancia de las y_i, y r_{xy} se le denomina coeficiente de correlación lineal. Este coeficiente es un índice que proporciona una idea de que tan agrupadas están las parejas de valores x_i, y_i a la curva de ajuste, en este caso a una línea recta. Obsérvese que si r_{xy} vale 1 ó - 1, de la ec 8.19 se obtiene que la variancia del error es cero, y por tanto, todos los puntos (X_i, Y_i) están sobre la curva o una recta.

Conforme el valor de r_{xy} tiende a cero, la correlación de los puntos en estudio se aleja de una línea recta. Si r_{xy} vale cero implica que la variancia del error es igual a la variancia de la variable dependiente y, y en este caso, la ecuación de regresión no es mejor que la media para estimar la variable dependiente y por lo tanto, no hay correlación entre las dos variables.

Si se analiza la ec 8.13 y 8.14, se ve que para cada valor inferido de la variable dependiente se tendrá un cierto error, en función de que tan correlacionadas estén las variables. Una medida de la variación de los puntos con respecto a la recta de regresión se puede deducir del error estándar de la estimación, que es análogo a la desviación estándar de una variable cuando se trata de conocer la dispersión respecto de su media. Para cada valor de la variable independiente $x = x_o$, se puede conocer cual es el error estándar para un cierto nivel de significancia α de la variable dependiente y, al utilizar la ec 8.13, aplicando la ecuación:

$$\mathcal{E} = \pm t_{\alpha/2} \; Se \sqrt{1 + \frac{1}{n} + \frac{n\,(xo - \bar{x})^2}{Sxx}} \qquad (8.20)$$

donde $t\alpha/_2$ se obtiene para la distribución "t" con n-2 grados de libertad.

De esta manera, de las ecs 8.13 y 8.20, se obtiene que

$$y = a + bx \pm \mathcal{E} \qquad (8.21)$$

que permite calcular el valor de la variable dependiente y con un cierto intervalo de confianza, para cualquier valor de la variable independiente x.

Ejemplo 8.7 Aplicación de la correlación lineal simple.

En la tabla mostrada (8.9), se proporcionan los gastos medios mensuales aforados por estaciones hidrométricas sobre una corriente localizada en el límite de los Estados de Tabasco y Chiapas dentro de la cuenca del Río Grijalva durante el año de 1969. Se desea obtener, por medio de un análisis de correlación lineal simple, la ecuación que relaciona dichos gastos y su coeficiénte de correlación, así como el gasto medio esperado en el Río Mezcalapa cuando se presente en el Río Grijalva un gasto medio de Q Med = 3,500 m^3/seg.

Como se trata de obtener una relación del tipo (ec 8.13): $y^1 = a + bx$, asignamos a los gastos las variables y^1 = variable dependiente a la Est. Reforma del Río Mezcalapa; x = variable independiente a la Est. Malpaso II sobre el Río Grijalva. Para facilidad de cálculos elaboramos la tabla 8.10.

Tabla 8.9 Gastos medios mensuales aforados durante el año de 1969

Orden	1969 Mes	Río Grijalva Estación Malpaso II Gasto medio (m^3/s)	Río Mezcalapa Estación Reforma Gasto medio (m^3/s)
1	Enero	321.08	175.97
2	Febrero	222.81	75.83
3	Marzo	155.41	45.94
4	Abril	274.58	77.57
5	Mayo	431.65	131.18
6	Junio	446.52	136.05
7	Julio	456.84	171.13
8	Agosto	1270.04	475.75
9	Septiembre	2089.29	897.42
10	Octubre	1618.41	710.58
11	Noviembre	431.72	268.30
12	Diciembre	509.33	224.12

Tabla 8.10 Ordenamiento del cálculo (ejemplo 8.7)

1 Orden	2 xi	3 yi	4 $(xi)^2$	5 $(yi)^2$	6 (xi)(yi)
1	321.08	175.97	103092.37	30965.44	56500.45
2	222.81	75.83	49644.30	5750.19	16895.68
3	155.41	45.94	24152.27	2110.48	7139.54
4	274.58	77.57	75394.18	6017.10	21299.17
5	431.65	131.18	186321.72	17208.19	56623.85
6	446.52	136.05	199380.60	18509.60	60749.05
7	456.84	171.13	208702.79	29285.48	78179.03
8	1270.04	475.75	1613001.60	226338.06	604221.53
9	2089.29	897.42	4365132.70	805362.66	1874970.63
10	1618.41	710.58	2619250.93	504923.94	1150009.78
11	431.72	268.30	186382.16	71984.89	115830.48
12	509.33	224.12	259417.05	50229.77	114151.04
	8227.68	3389.84	9889872.18	1768685.80	4156570.23

De la Tabla 8.10 se obtienen los siguientes valores:

$$\sum_{i=1}^{n} X_i = 8227.68$$

$$\sum_{i=1}^{n} Y_i = 3389.84$$

$$\sum_{i=1}^{n} (X_i)^2 = 9,889,872.18$$

$$\sum_{i=1}^{n} (Y_i)^2 = 1,768,685.80$$

$$\sum_{i=1}^{n} X_iY_i = 4,156,570.23$$

Cálculo de los parámetros a y b

De la ec (8.17) se deduce que

$$Sxx = 12\ (9889,872.18) - (8227.68)^2 = 50983747.96$$

y de la ec (8.18) $Sxy = 12\ (4156570.23) - (8227.68)\ (3389.84) = 21988323.9$

La media de los datos (ec 8.6) son:

$$\bar{y} = \frac{3389.84}{12} = 282.48$$

$$\bar{x} = \frac{8227.68}{12} = 685.64$$

Sustituyendo los valores anteriores en las ecs (8.15) y (8.16) se obtiene

$$b = \frac{21988323.99}{50983747.98} = 0.4310$$

$$a = 282.48 - 0.4310\ (685.64) = -13.03$$

Por tanto, la ecuación de la recta de regresión que proporciona el mejor ajuste entre los valores de los gastos medios mensuales en los ríos Grijalva (x) y Mezcalapa (y) resulta según la ec. 8.13

$$y' = -13.03 + 0.4310\ x$$

siendo su coeficiente de correlación, teniendo en cuenta que

$$Syy = 12\ (1768685.80) - (3389.84)^2 = 9733214.38$$

igual a

$$r^2 = \frac{(21988323.99)^2}{(50983747.98)\ (9733214.38)}\ ;\ r = 0.9871$$

En la fig. 8.4 se muestran los valores datos, así como la ecuación de mejor ajuste entre ellos.

Para conocer el gasto medio que circula por el Río Mezcalapa, cuando aguas abajo en el Río Grijalva se presentó un gasto de 3500 m^3/seg de la ec. de la recta de regresión encontrada se tiene que

$$Q\ \text{Mezcalapa} = f\ (Q\ \text{Grijalva})$$

$$y = f\ (x)$$

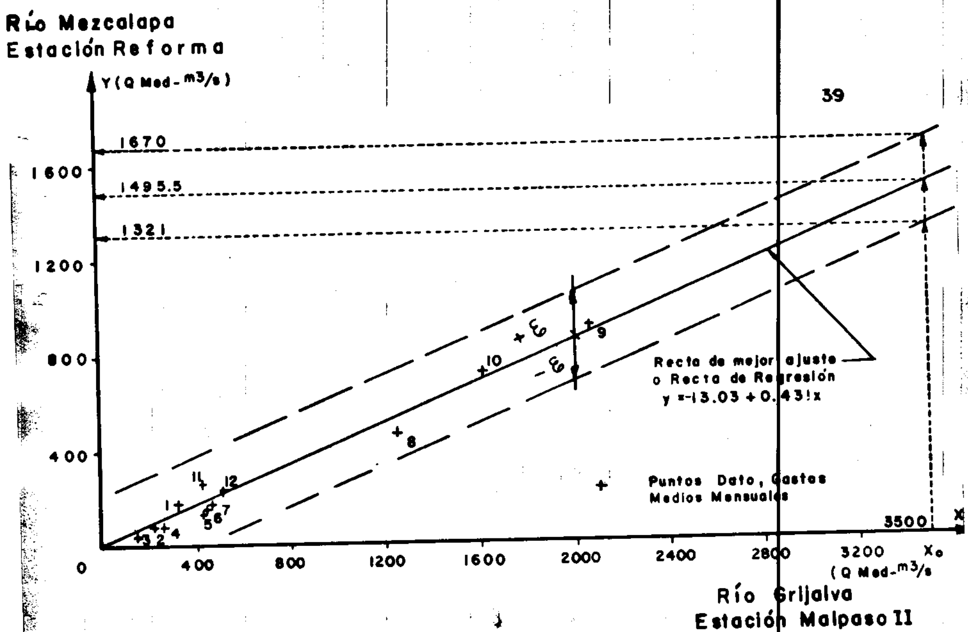


Fig. 8.4 Recta de Regresión obtenida para relacionar los gastos medios mensuales, aforados en las estaciones hidrométricas Reforma y Malpaso II.

Por lo que sí $x = 3500$

$$y' = -13.03 + 0.4310\,(3500) = 13.03 + 1508.50$$

$$y' = 1495.47 \quad m^3/seg.$$

resultaría el gasto medio en el Río Mezcalpa, obtenido de la recta de regresión.

Considerando para valuar el error estándar de la predicción un nivel de significancia de 0.95, ($\alpha = 0.05$) y $\alpha/2 = 0.025$ y grados de libertad

$$\nu = n - 2 = 12 - 2 = \underline{10}$$

de los valores tabulados de la distribución "t" se deduce:

$$t_{\frac{\alpha}{2}} = t_{0.025} = \underline{\underline{2.228}}$$

siendo la variancia del error (ec 8.19) igual a

$$Se^2 = \frac{Syy}{n(n-2)} \left(1 - \frac{(Sxy)^2}{SxxSyy}\right) = \frac{9733214.38}{12\,(12-2)} \left(1 - \frac{(21988323.99)^2}{(50983747.98)(9733214.38)}\right)$$

$$= 81110.12\,(1 - 0.97) = 8110.12\,(0.03) = 2083.89$$

$$Se = \sqrt{2083.89} = \underline{\underline{45.65}}$$

Sustituyendo los valores en la ec 8.20, con $Xo = 3500$

$$\mathcal{E} = \pm\,(2.228)\,(45.65)\sqrt{1 + \frac{1}{12} + \frac{12\,(3500 - 685.64)^2}{50983747.98}} =$$

$$= \pm\,101.7082\,\sqrt{1 + 0.0833 + 1.8643} = \pm\,101.7082\,(1.716)$$

así, el intervalo de confianza: $\mathcal{E} = \pm\;174.62$

por tanto, se podría afirmar que el gasto medio más probable $(P = 0.95)$ esperado en la estación Reforma cuando se presentan $3500\ m^3/seg.$ en la estación Malpaso II sería, con su intervalo de confianza de.

$$y = 13.03 + 0.4310\,(3500) \pm 174.62$$

$$y = 1495.47 \pm 174.62$$

así

$$Q\ Med\ máximo = 1495.47 + 174.62 = 1670\ m^3/seg$$

$$Q\ Med\ mínimo = 1495.47 - 174.62 = 1321\ m^3/seg$$

bad

Puede notarse que el error estándar de la predicción Y_o intervalo de confianza resulta el 11.7 % del valor obtenido para y'.

Ejemplo 8.8 En la tabla (8.11) se proponen los gastos máximos mensuales aforados (col. 3) y su correspondiente lectura sobre la escala de referencia (col. 4) del Río Mezcalapa, Estación Las Peñitas, en el Edo. de Chiapas correspondiente a la cuenca del Río Grijalva. El cero de la escala se encuentra en la cota + 50.00 m respecto al banco de nivel. Se desea encontrar, aplicado el método de correlación lineal simple, la ecuación que mejor relacione dichos valores, su coeficiente de correlación, así como el gasto máximo mensual esperado en la corriente cuendo se tenga una lectura de escala de 6.80 m.

La relación pedida corresponde que la ecuación de una curva de elevaciones–gastos para el río en estudio. Esta curva generalmente tiene la forma :

$$Q = a_o h^b$$

Q gasto que pasa por sección (m^3/seg)

h tirante hidráulico o elevación de la superficie del agua respecto a un punto. (m)

a_o,b parámetros, función de las características particulares del escurrimiento.

Siendo esta la forma de la ecuación buscada, para poder utilizar el análisis de correlación lineal en vez de la curvilínea, será necesario efectuar ciertas transformaciones de forma tal que pueda relacionar linealmente, así tomando logarítmos se tiene

$$\log Q = \log a_o + b \log h$$

y considerando

$$y' = \log Q$$

$$x = \log h$$

$$a = \log a_o$$

queda

$$y' = a + bx$$

que es la ecuación de una recta con pendiente b y ordenada al origen a, con la característica de la recta de regresión pudiendo así emplear el análisis lineal deseado. Para la secuencia de cálculo, se elaboró la Tabla 8.12

De dicha tabla se tienen los siguientes valores:

$$\sum_{i=1}^{n} x_i = 7.0620 \quad ; \quad \sum_{i=1}^{n} y_i = 36.1844$$

$$\sum_{i=1}^{n} (x_i)^2 = 4.3971 \quad ; \quad \sum_{i=1}^{n} (y_i)^2 = 110.7134$$

Tabla 8.11 Gastos máximos mensuales en la escala de tirantes correspondiente al Río Mezcalapa en la estación La Peñita durante el año de 1969.

1	2	3	4
Orden	Mes	Gasto máximo mensual (m^3/s)	Lectura en la escala de tirantes respecto al cero (m)
1	Enero	1156.000	4.16
2	Febrero	445.740	2.75
3	Marzo	325.450	2.41
4	Abril	457.000	2.86
5	Mayo	506.500	2.97
6	Junio	560.500	3.09
7	Julio	884.500	3.69
8	Agosto	2230.000	5.30
9	Septiembre	4000.000	6.40
10	Octubre	3765.000	6.42
11	Noviembre	2490.000	5.39
12	Diciembre	944.000	3.68

Tabla 8.12 Ordenamiento de cálculo.

1	2	3	4	5	6	7	8
Orden	h	q	xi	yi	$(xi)^2$	$(yi)^2$	(xi)(yi)
1	4.16	1156.00	0.6191	3.0645	0.3832	9.3911	1.8972
2	2.75	445.74	0.4393	2.6493	0.1929	7.0187	1.1638
3	2.41	325.45	0.3820	2.5119	0.1459	6.3096	0.9595
4	2.86	457.00	0.4564	2.6599	0.2083	7.0750	1.2139
5	2.97	506.50	0.4728	2.7050	0.2235	7.3170	1.2789
6	3.09	560.50	0.4900	2.7490	0.2401	7.5570	1.3470
7	3.69	884.50	0.5670	2.9469	0.3214	8.6842	1.6708
8	5.30	2230.00	0.7243	3.3483	0.5246	11.2111	2.4251
9	6.40	4000.00	0.8062	3.3483	0.6499	12.9751	2.9040
10	6.42	3765.00	0.8075	3.5763	0.6520	12.7899	2.8878
11	5.39	2490.00	0.7316	3.3962	0.5352	11.5341	2.4846
12	3.68	944.00	0.5658	2.9750	0.3201	8.8586	1.6832
		Suma:	7.0620	36.1844	4.3971	110.7134	21.9158

$$\sum_{i=1}^{n} x_i y_i' = 21.9158$$

sustituyendo los valores anteriores en la ec (8.17) se tiene que

$$Sxx = 12(4.3971) - (7.0620)^2 = 2.8934$$

y de la ec (8.18)

$$Sxy = 12(21.9158)(36.1844) = 7.4554$$

siendo la medida de los datos:

$$\bar{y} = \frac{36.1844}{12} = 3.0153$$

$$\bar{x} = \frac{7.0620}{12} = 0.5885$$

Sustituyendo los valores en las ecs (8.15) y (8.16) se tiene que

$$b = \frac{7.4554}{2.8934} = 2.58 \quad y$$

$$a = 3.0153 - 2.58\,(0.5885) = 1.50$$

Por tanto, la ecuación de la recta de regresión (en escala doble-logarítmico por la transformación hecha) que proporciona el mejor ajuste entre los valores de los gastos máximos mensuales (y) y su correspondiente lectura en la escala (x) sobre el Río Mezcalapa resulta, según la ec.(8.13)

$$y' = 1.50 + 2.58\ x$$

pero de acuerdo con la transformación hecha y la forma de la ecuación de una curva E-Q, se tiene que

$$a = \log a_o = 1.50$$

$$a_o = \text{antilog } 1.50 = 31.62$$

así la ecuación de la cifra E-Q que proporciona el mejor ajuste a los datos en escala normal o aritmética resulta

$$Q = 31.62\,(h)^{2.58}$$

En las figs 8.5 y 8.6 se muestran las relaciones existentes entre los datos y estas ecuaciones.

Su coeficiente de correlación considerando

$$Syy = 12(110.7134) - (36.1844)^2 = 19.25$$

se obtiene de la ec (8.19) como

$$r = \frac{7.4554}{\left[(2.8934)(19.2500)\right]^{1/2}} = \frac{7.4554}{7.4630}$$

$$r = 0.9989$$

que involucra un ajuste practicamente sin error apreciable.

Para conocer el gasto máximo mensual que circula por la sección en el Río Mezcalapa cuando se tiene una lectura de 6.80 m en la escala aforadora en su margen, se pueden utilizar las ecuaciones de la recta de regresión y de la curva encontradas. Así,

$$Q \text{ Mezcalapa} = f \text{ (h escala)}$$

$$y = f\ (x)$$

como:

$$y' = 1.50 + 2.58\ x$$

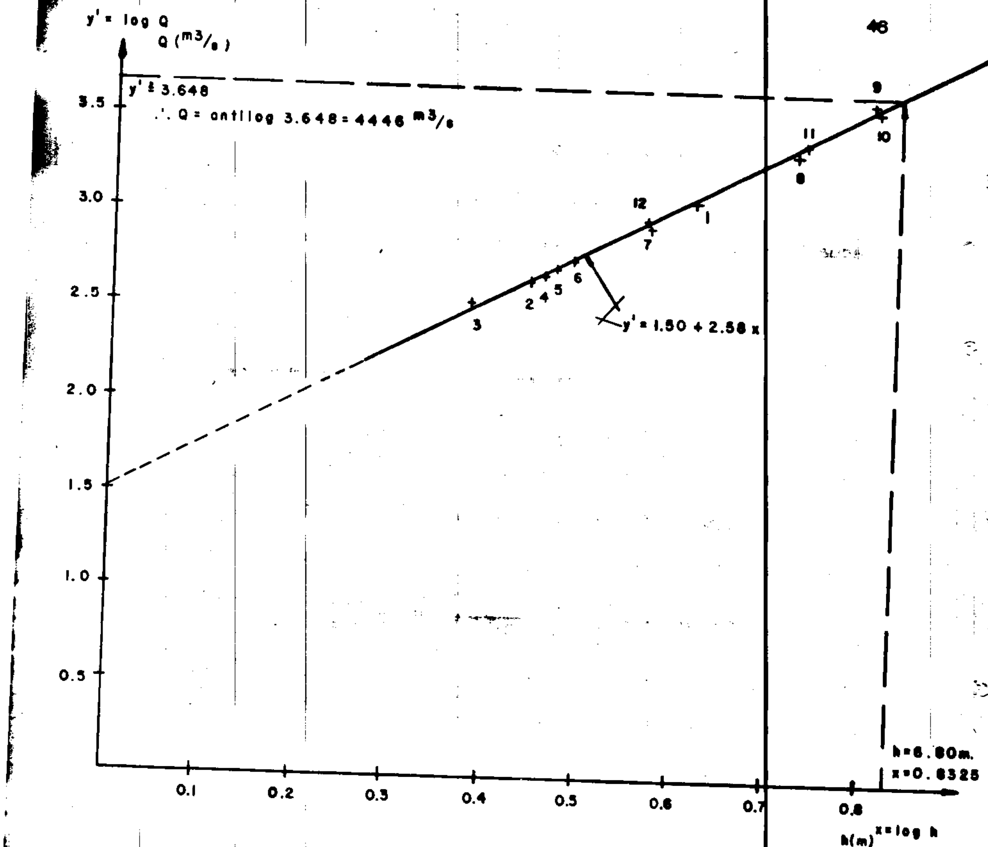


Fig. 8.5 Recta de Regresión ($y = 1.50 + 2.58x$) obtenida para relacionar los gastos máximos mensuales en el Río Mezcalapa y las lecturas correspondientes en la escala de Referencia. Escala aritmética.

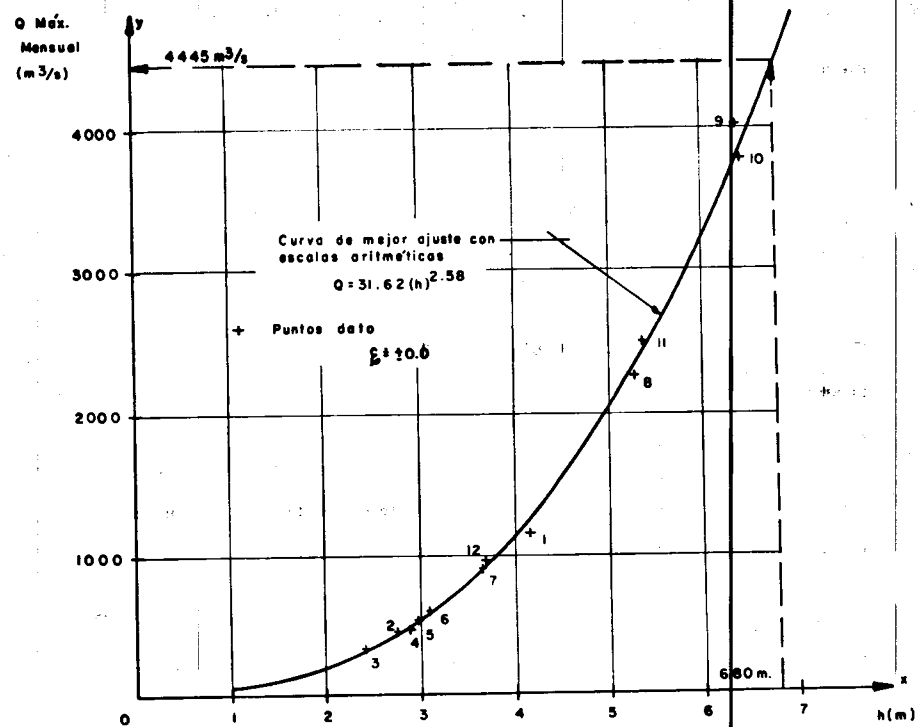


Fig. 8.6 Curva de mejor ajuste obtenida para relacionar los gastos máximos mensuales en el Río Mezcalapa y las lecturas correspondientes en la escala de referencia. Curva Elevaciones-Gastos de la Estación Hidrométrica Las Peñitas.

siendo $x = \log h = \log 6.80 = 0.8325$

subst: $y' = 1.50 + 2.58\,(0.8325) = 1.50 + 2.1478 = 3.6478$

$y = \log Q;\ Q = \text{antilog}\,(y') = \text{antilog}\,(3.6478)$

$QM = \underline{4444}\ m^3/seg$ **recta de regresión , o bien la** ecuación de la curva E - Q.

$$Q = 31.62\,(h)^{2.58}$$

$$= 31.62\,(6.80)^{2.58} = 31.62\,(140.60) = \underline{4445.80}\ m^3/seg$$

Si se elige para valuar el error un nivel de significancia de 0.95, $\alpha = 0.5$

y $\frac{\alpha}{2} = 0.025$

los grados de libertad $\nu = n-2 = 12-2 = 10$, **de los valores tabulados para** la distribución "t" obtenemos:

$$t_{\frac{\alpha}{2}} = t_{0.025} = 2.228 \quad y$$

la variancia del error "Se^2" de acuerdo a la ec (8.19) es

$$Se^2 = \frac{19.25}{12(12-2)}\left[1 - \frac{(7.4554)^2}{(2.8934)(19.25)}\right] = 0.00034$$

y la desviación estándar: $Se = 0.018$

Sustituyendo los valores en la ec (8.20) con $Xo = 0.8325$ se tiene

$$\varepsilon = \pm\,(2.228)\,(0.018)\sqrt{1 + \frac{1}{12} + \frac{12\,(0.8325 - 0.5885)^2}{2.8934}} =$$

$$\varepsilon = \pm\ 0.0401 \sqrt{1 + 0.0833 + 0.2467} = \pm\ 0.0401 \sqrt{1.33} =$$

$$= \pm\ 0.0401\ (1.153) = \pm\ 0.0462$$

$$= \pm\ 0.0462$$

por tanto, de la ec (8.21)

$$y = a + bx \pm \varepsilon$$

vemos que el valor para el error estándar o intervalo de confianza - $\varepsilon = \pm\ 0.0462$ en esta curva elevaciones - gastos resulta tan pequeño, que el valor del gasto máximo mensual más probable que se presenta en el Río Mezcalapa cuando la lectura sobre la escala es de 6.80 m, se obtiene directamente de las ecuaciones encontradas, depreciando el error o intervalo de confianza, sin que esto afecte a la predicción en forma - significativa, así

cuando $h = 6.80$ m

$QM = 4445\ m^3/seg$

Lo anterior es de esperarse pués el coeficiente de correlación prácticamente tiene el valor de la unidad.

8.3 Correlación lineal múltiple

Esta técnica de análisis se utiliza cuando la variable dependiente "y" es función de dos o más variables independientes $x_1\ x_2 \ldots x_n$. Es muy usada en hidrología para obtener relaciones por ejemplo entre los gastos máximos y las características fisiográficas de la cuenca en estudio, para determinar fórmulas de tiempos de pico, para generación de escurrimientos, etc. El valor por estimar y' se puede conocer a partir de una ecuación lineal del tipo

$$y' = a_o + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \qquad (8.22)$$

donde las a_i se determinan a partir de los datos disponibles y de tal manera que la suma de los errores al cuadrado sea mínima. A partir de esto, los parámetros a_i se obtienen al resolver el siguiente sistema:

$$\begin{pmatrix} N & x_{1i} & x_{2i} & x_{ni} \\ x_{1i} & x_{1i}^2 & x_{1i}\,x_{2i} & x_{1i}\,x_{ni} \\ & & & \\ x_{ni} & x_{ni}\,x_{1i} & x_{ni}\,x_{2i} & x_{ni}^2 \end{pmatrix} \begin{pmatrix} a_o \\ a_1 \\ \\ a_n \end{pmatrix} = \begin{pmatrix} y_i \\ x_{1i}\,y_i \\ \\ x_{ni}\,y_i \end{pmatrix} \qquad (8.23)$$

donde N es el número de grupos de valores ($x_{1i}, x_{2i}, \ldots x_{ni}, y_i$) y las sumas son desde i hasta N.

A partir de una ecuación similar a la (8.14) y de acuerdo con la ec (8.9) se puede obtener la variancia del error como

$$Se^2 = \sum y_i^2 - s_n^2 \qquad (8.24)$$

donde

$$S_n^2 = a_o \sum y_i + a_1 \sum x_{1i} y_i + \ldots + a_n \sum x_{ni} y_i \qquad (8.25)$$

Las ecs (8.24) y (8.25) permiten conocer que tanto influyen cada una de las variables independientes (x_1 , x_2 , ... x_n) en el valor de y. Suponiendo que se desea saber que tanto influye la variable x_n en el cálculo de y, se procede de la siguiente manera; se calcula utilizando la ec (8.25) la variacia S_{n-1}^2 de las restantes variables, pero sin tener en cuenta el último término función de x_n. Conocidas las variancias, se aplica la - prueba F (ec 8.10) considerando que

$$F = \frac{N-n-1}{Se^2} \quad \frac{(Sn^2 - Sn^2-1)}{n-(n-1)} \qquad (8.26)$$

Comparando el valor calculado de F según la ec (8.26) y el valor tabular de F con grados de libertad $\nu_1 = N-n-1$ y $\nu_2 = N-(n-1)$, si el primero es mayor que él segundo, el ajuste de"y"mejora si se utiliza x_n.

Obsérvese que el criterio anterior se puede efectuar suponiendo simultáneamente varias variables x_1, pero siempre comparando con respecto a la ec (8.22). Por el proceso iterativo de este análisis, conviene - siempre empezar comprando el valor observado de "y"con el obtenido - mediante el ajuste de"y," este último función de dos y tres variables de pendientes, y así sucesivamente.

Otra manera de ver como influye cada variable de x_i. Para cada ec (8.22) se valúan los parámetros a_i y se calculan con los valores datos los valores de y', obteniendose finalmente el coeficiente de correlación entre estos valores y los valores datos de y. De esta manera se encuentra cual es la mejor ec (8.22) para determinar los valores de y.

Ejemplo 8.9 A partir del análisis de las características de las subcuencas aforas más representativas correspondientes a la cuenca del Río Pánuco, se desea obtener una relación que permita conocer el gasto máximo para diversos períodos de retorno. Una vez definida la relación, obtener el gasto máximo para la cuenca del Río Pánuco hasta su desembocadura para un período de retorno de 25 años.

Las cuencas que se seleccionaron para el análisis se muestran en la Tabla 8.13. La información de estaciones, gastos máximos y año de registros se obtuvieron del "Boletín Hidrológico No. 32, Tomo I" editado por la Secretaría de Recursos Hidráulicos, a las cuales se les calculó sus características fisiográficas más importantes (capítulo 2), relacionando estas con el gasto máximo registrado y su período de retorno (véase inciso 8.4.3)

TABLA 8.13

CARACTERISTICAS DE LAS SUBCUENCAS SELECCIO-NADAS PARA EL ANALISIS

Corriente Río	Estación Nombre	Area (km^2) A	Pendiente cauce principal S	Long. Cauce Principal (m) L
Tampaón	El Pujal	23373	0.0076	325 000
Moctezuma	Pte. Mazacintla	17238	0.0077	215 000
Amajac	Temamatla	6884	0.0135	155 000
Tempoal	Tempoal	5275	0.0139	150 000
Axtla	Requetemú	661	0.0792	25 000

Corriente Río	Estación Nombre	Desnivel entre origen y sección de aforos (m) E	Gasto máximo (m^3/seg) $Q_{máx}$	Período de retorno (años) Tr
Tampaón	El Pujal	2474	5410	6.50
Moctezuma	Pte. Mazacintla	1647	5233	4.33
Amajac	Temamatla	2087	4037	3.25
Tempoal	Tempoal	2085	3295	2.60
Axtla	Requetemú	1979	2696	2.17

Después de revisar diversas ecuaciones, se propuso para iniciar el estudio una ecuación de liga del tipo

$$Q_M = a\ A^b S^c T^d$$

que involucra la relación basada, siendo a, b, c y d los parámetros a determinar. Aunque lógicamente al hacer intervenir en el problema solo cinco subcuencas, para procesar la ecuación planteada, se tiene muy poca información para realizar el ajuste de los parámetros y utilizarlos para hacer inferencia. Por claridad se prefirió sacrificar información y trabajar solo con cinco subcuencas. Hecha la aclaración anterior, como la ecuación propuesta representa una relación no lineal, se hace la transformación necesaría para aplicar el procedimiento visto en la correlación líneal multiple. Tomando logarítmos decimales se tiene que

$$\log Q_M = \log a + b \log A + c \log S + d \log Tr$$

$$\log Q_M = y'$$

con lo cual

$$\log a = a_o,\ b = a_1, c = a_2,\ d = a_3$$

$$\log A = x_1,\ \log S = x_2,\ \log Tr = x_3$$

queda $y' = a_o + a_1 x_1 + a_2 x_2 + a_3 x_3$, que es la relación lineal pedida. De acuerdo al sistema (8.23) indicando con

$$\sum_{i=1}^{n} = \Sigma \ ; x_{1i} = x_1, \ x_{2i} = x_2, \ x_{3i} = x_3 \ , \ y_i = y,$$

las ecuaciones asociadas a la matríz serían

$$y = a_o n + a_1 \ \Sigma \ x_1 + a_2 \ \Sigma \ x_2 + a_3 \ \Sigma \ x_3$$

$$yx_1 = a_o \ \Sigma \ x_1 + a_1 \ \Sigma \ x_1^2 + a_2 \ \Sigma \ x_1 x_2 + a_3 \ \Sigma \ x_1 x_3$$

$$yx_2 = a_o \ \Sigma \ x_2 + a_1 \ \Sigma \ x_1 x_2 + a_2 \ \Sigma \ x_2^2 + a_3 \ \Sigma \ x_2 x_3$$

$$yx_3 = a_o \ \Sigma \ x_3 + a_1 \ \Sigma \ x_1 x_3 + a_2 \ \Sigma \ x_2 x_3 + a_3 \ \Sigma \ x_3^2 \ldots \text{(I)}$$

que es un sistema de 4 ecuacones con 4 incógnitas. Para resolver el sistema se elaboró la Tabla 8.14

Substituyendo los resultados de la Tabla 8.14 en el sistema, se tiene que

$$18.008 = 5a_o + 18.986 \ a_1 - 9.060 \ a_2 + 2.714 \ a_3$$

$$68.681 = 18.986 \ a_o + 73.576 \ a_1 - 35.404 \ a_2 + 10.712 \ a_3$$

$$-32.824 = 9.060 \ a_o - 35.404 \ a_1 + 17.112 \ a_2 - 5.168 \ a_3$$

$$9.868 = 2.714 \ a_o + 10.712 \ a_1 - 5.168 \ a_2 + 1.615 \ a_3$$

Resolviendo el sistema, se deduce los valores de los parámetros que son:

$a_o = 5.5541$	=	log a	a = antilog a_o = 358,179
$a_1 = 1.8736$	=	b	
$a_2 = 2.2136$	=	c	
$a_3 = 2.1207$	=	d	

Tabla 8.14 Ordenamiento del cálculo

Orden	Q máx (m^3/s)	Area ($/2m^2$)	Pendiente (S^{-m}/m)	Período de Retorno (Tr-años)	y log Q	x_1 log A	x_2 log S	x_3 log Tr
1	5410	23373	0.0076	6.50	3.735	4.369	-2.119	0.813
2	5233	17238	0.0077	4.33	3.720	4.238	-2.113	0.637
3	4037	6884	0.0135	3.25	3.606	3.837	-1.870	0.512
4	3295	5275	0.0139	2.60	3.516	3.516	-1.857	3.722
5	2696	661	0.0792	2.17	3.431	2.821	-1.101	0.337
					18.008	18.986	-9.060	2.714

x_1^2	x_2^2	x_3^2	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$y x_1$	$y x_2$	$y x_3$
19.088	4.490	0.661	-9.258	3.552	-1.723	16.318	-7.914	3.037
17.961	4.465	0.406	-8.955	2.700	-1.346	15.765	-7.860	2.370
14.723	3.497	0.262	-7.175	1.965	-0.957	13.836	-6.743	1.846
13.846	3.448	0.172	-6.910	1.544	-0.771	13.083	-6.529	1.459
7.958	1.212	0.114	-3.106	0.951	-0.371	9.679	-3.778	1.156
73.576	17.112	1.615	-35.404	10.712	-5.168	68.681	-32.824	9.868

y finalmente la ecuación buscada resulta

$$y' = a_o + a_1x_1 + a_2x_2 + a_3x_3$$

$$\log Q_M = \log a + b \log A + c \log S + d \log Tr$$

$$\log Q_M = 5.5541 - 1.8736 \log A - 2.2136 \log S + 2.1207$$

$$\log Tr \ . \ . \ . \ (I)$$

$$Q_M = a\, A^b\, S^c\, Tr^o$$

así

$$Q_M = 358179 \quad A^{-1.8736}\, S^{-2.2136}\, Tr^{\,2.1207}$$

$$Q_m = 358179 \ \frac{Tr^{2.12}}{(A^{1.87})(S^{2.21})} \ (m^3/seg)$$

y de acuerdo a la ecuación (8.24) la variancia del error será:

$$Se^2 = \quad y_i^2 - S_n^2$$

donde (8.25) $\quad S_n^2 = a_o \sum y + a_1 \sum yx_1 + a_2 \sum yx_2 + a_3 \sum yx_3 =$

$$= 5.5541\ (18.008) - 1.8736\ (68.681) - 2.2136$$

$$(-32.824) + 2.1207\ (9.868) = 64.9238$$

por otro lado

$$y_i^2 = 64.9259$$

subst en (8.24)

$$Se^2 = 64.9259 - 649238 = 0.0021$$

y la desv. estándar del error.

$$Se = \sqrt{Se^2} = \sqrt{0.0021} = 0.0456$$

S el gasto máximo para el Río Pánuco asociado a un Tr = 25 años, resultaría según la ecuación obtenida, en la estación de Las Adjuntas - A = 61063 km^2

$$\log Q_M = a_o + a_1x_1 + a_2x_2 + a_3x_3 \ ; \ A = 61063 \ km^2; \ S = 0.0071; \ Tr = 25 \text{ años}$$

$$= 5.5541 - 1.8736 (\log A) - 2.2136 (\log S) + 2.1207 (\log Tr)$$

$$= 5.5541 - 1.8736 (4.7858) - 2.2136 (-2.1487) + 2.1207 (1.3979)$$

$$= 4.3085 \qquad \log Q_M = 4.3085$$

$$Q_M = \text{antilog } 4.3085$$

$$Q_M = 20340 \ m^3/seg$$

$$Tr = 25 \text{ años}$$

Ejemplo 8.10 Se desea conocer la relación existente entre los gastos máximos (Qmáx) del Río Pánuco en función de sus afluentes principales como lo son los Ríos Moctezuma, Amajac y Tempoal. La ecuación pedida resultaría del tipo (Q = Qmáx)

$QP = a + bQM + c\,QA + d\,QT$, y obtener los registros de 1959 y 60 de la est. Temamatla faltantes.

Para obtener la relación deseada, se eligiaron sobre las corrientes afectadas, las estaciones hidrométricas más representativas en cuanto a sus aforos de gastos circulantes, resumiendo estas:

(P) Río Pánuco - Estación Las Adjuntas

(M) Río Moctezuma - Estación Tierra Blanca

(A) Río Amajac - Estación Temamatla

(T) Río Tempoal - Estación Tempoal, asignando a cada una de ellas un período de registro de 8 años para el proceso de cálculo, datos que se muestran en la Tabla 8.15

siendo la relación entre los gasto máximos:

$$QP = a + b\,QM + c\,QA + d\,QT$$

$$QP = y,\ QM = x_1\ ,\ QA = x_2\ ,\ QT = x_3$$

$$a = a_0,\ b = a_1,\ c = a_2\ ,\ d = a_3 \text{ queda:}$$

$$y = a_0 + a_1\,x_1 + a_2\,x_2 + a_3\,x_3$$

relación que es posible resolverla, encontrando los valores de los parámetros a_0, a_1, a_2 y a_3 tales que proporcionen el mejor ajuste a los valores reales y por medio del análisis de correlación lineal múltiple - (con 4 valores); así de acuerdo al sistema (8.23) el planteamiento material se transforma indicando con:

$$\sum_{i=1}^{n} = \sum\ ;\ x_{1i} = x_1,\ x_{2i} = x_2,\ x_{3i} = x_3,\ y_i = y$$

a las ecuaciones

$$\Sigma y = a_{o}n + a_1 \Sigma x_1 + a_2 \Sigma x_2 + a_3 \Sigma x_3$$
$$\Sigma yx_1 = a_o \Sigma x_1 + a_1 \Sigma x_1^2 + a_2 \Sigma x_1x_2 + a_3 \Sigma x_1x_3$$
$$\Sigma yx_2 = a_o \Sigma x_2 + a_1 \Sigma x_1x_2 + a_2 \Sigma x_2^2 + a_3 \Sigma x_2x_3$$
$$yx_3 + a_o \Sigma x_3 + a_1 \Sigma x_1x_3 + a_2 \Sigma x_2x_3 + a_3 \Sigma x_3^2$$

que es un sistema de 4 ecuaciones con 4 incógnitas. Para resolver el sistema se elaboró la Tabla 8.16

Sustituyendo los valores de la Tabla 8.16 en el sistema se tiene que

$$27.065 = 8a_o + 4.302a_1 + 6.616a_2 + 10.101a_3$$
$$17.8451 = 4.302a_o + 4.9336a_1 + 4.2504a_2 + 7.182a_3$$
$$22.5447 = 6.616a_o + 4.2504a_1 + 5.9343a_2 + 8.5987a_3$$
$$39.4159 = 10.101a_o + 7.1820a_1 + 8.5987a_2 + 15.2140a_3$$

Resolviendo el sistema se encuentran las valores de los parámetros que son $a_o = 1.2496 = a$; $a_1 = -0.0164 = b$; $a_2 = -0.8029 = c$; $a_3 = 2.2226 = d$

y la ecuación buscada es

$$y = a_o + a_1x_1 + a_2x_2 + a_3x_3$$
$$QP = a + b\,QM + c\,QA + d\,QT$$
$$QP = 1.2496 - 0.0164\,QM - 0.8029\,QA + 2.2226\,QT$$
$$\text{en } (m^3 \times 10^3) \qquad \ldots(I)$$

De acuerdo a la ecuación (8.24) la variancia del error será:

$$Se^2 = \Sigma\ y_i^2 - S_n^2$$

Tabla 8.15 Relación entre gastos máximos aforados.

Gastos máximos aforados en miles de m^3
Estación

Orden	Año	Las Adjuntas y	Tierra Blanca x_1	Tamamatla x_2	Tempoal x_3
1	1961	3.295	0.290	0.828	0.853
2	1962	1.735	0.157	0.642	0.739
3	1963	4.037	0.287	0.774	1.800
4	1964	2.038	0.225	0.604	0.748
5	1965	2.621	0.327	0.856	0.793
6	1966	5.410	0.341	0.522	1.778
7	1967	5.233	0.005	1.118	2.245
8	1968	2.696	0.670	1.272	1.145
n=8	SUMA	27.065	4.302	6.616	10.101

Tabla 8.16 Procesamiento de cálculo.

x_1^2	x_2^2	x_3^2	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	yx_1	yx_2	yx_3
0.0841	0.6855	0.7276	0.2401	0.2473	0.7062	0.9555	2.7282	2.8106
0.0246	0.4121	0.5461	0.1007	0.1160	0.4744	0.2723	1.1138	1.2821
0.0823	0.5990	3.2400	0.2221	0.5166	1.3932	1.1586	3.1246	7.2666
0.0506	0.3648	0.5595	0.1359	0.1683	0.4517	0.4585	1.2309	1.5244
0.1069	0.7327	0.6288	0.2799	0.2593	0.6788	0.8570	2.2435	2.0784
0.1162	0.2724	3.1612	0.1780	0.6062	0.9281	1.8448	2.8240	9.6189
4.0200	1.2499	5.040	2.2415	4.5042	2.5099	10.4921	5.8504	11.7480
0.4489	1.6179	1.311	0.8522	0.7671	1.4564	1.8063	3.4293	3.0869
4.9336	5.9343	15.2142	4.2504	4.2504	7.180	17.8451	22.5447	39.4159

donde (8.25) $S_n^2 = a_o \Sigma\ y + a_1 \Sigma\ yx_1 + a_2 \Sigma\ yx_2 + a_3 \Sigma\ yx_3 =$

$= 1.2496\ (27.065) - 0.0164\ (17.8451) +$

$- 0.8029\ (22.5447) + 2.2226\ (39.4159) = 103.0324$

siendo $y_t^2 = 105.1085$

subst en (8.24)

$$Se^2 = 105.1085 - 103.0324 = 2.0761$$

y la desv. estándar del error

$$Se = \sqrt{Se^2} = \sqrt{2.0761} = \pm 1.441 \times 10^3\ m^3/seg$$

Los registros correspondientes a la estación Temamatla serían en función de la relación encontrada y las demás estaciones:

$$QA = \frac{1}{d} \left[QP - (a + b\ QM + QT) \right]$$

Los registros para 1959 y 1960 para las estaciones independientes son:

$Q_{máx}$ en ($m^3 \times 10^3$)

Año	Las Adjuntas (QP)	Tierra Blanca (QM)	Tempoal (QT)
1959	2.222	0.4310	1.508
1960	2.189	0.1741	1.277

subst para 1959:

$$QA_{59} = \frac{1}{2.2226} \left[2.222 - (1.2496 - 0.0164 \times 0.4310 - 0.8029 \times 1.508) \right] =$$

$$QA_{59} = 0.9854 \times 10^3\ m^3\ b = 985.4\ m^3/seg$$

para 1960 :

$$QA_{60} = \frac{1}{2.2226}\left[2.189 - (1.2496 - 0.0164 \times 0.1741 - 0.8029 \times 1.277)\right] =$$

$$QA_{60} = 0.8853 \times 10^{3} \text{ m/seg.} = 885.3 \text{ m}^{3}\text{/seg.}$$

8.4 Ordenamiento de datos hidrológicos

8.4.1 Selección de un registro

Los datos hidrológicos disponibles son generalmente presentados en orden cronológico. En la fig. 8.7 se muestra un registro de ocho años de gastos medios diarios en orden de ocurrencia incluyendo solo los picos de aquellos valores que son mayores o iguales a 250 m^3/seg.

En general muchos de los datos originales de que se disponen, no tienen significancia práctica en el análisis de los mismos, debido a que usualmente el diseño hidrológico de un proyecto, es gobernado solo por alguna ó algunas condiciones críticas. Por esta razón generalmente se trabaja solo con dos tipos de datos, conocidos como series de valores extremos y series de duración parcial.

La serie de valores extremos incluye solo el valor más grande ó más pequeño de los valores comprendidos en el registro, en un cierto intervalo constante de tiempo. Si el intervalo de tiempo es un año, la serie obtenida se le conoce como "serie anual", y si contiene los valores más grandes se le denomina serie de "máximos anuales"; si se refiere a los valores menores, se le llama serie de "mínimos anuales". (fig 8.7 b)

La serie de duración parcial constituye una serie de datos, los cuales se seleccionan de forma tal que su magnitud sea mayor que un cierto valor base, esto se escoge para que el número de valores en la serie sea igual al número de años de registro, la serie de datos resultante se le denomina "serie de excedentes anuales", (fig 8.7 c).

La selección de que tipo de serie de datos debe usarse en un diseño, se basa en la aplicación que se le va a dar al análisis, de los datos, los excedentes anuales se emplean si el segundo valor más grande en el año puede afectar en el diseño. La serie de máximas anuales se utiliza cuando en el diseño gobiernan las condiciones más críticas. En general estos tipos no difieren mucho, excepto en los valores bajos, y para efectos de comparación, conviene utilizar las dos series. Tradicionalmente los gastos se analizan como serie de máximos anuales y los lluviosos como serie de excedentes anuales.

8.4.2 Período de retorno

En un análisis de frecuencias de datos hidrológicos, el primer objetivo es determinar el intervalo de recurrencia o período de retorno T, de un evento hidrológico de una magnitud dada "y". El período de retorno se -

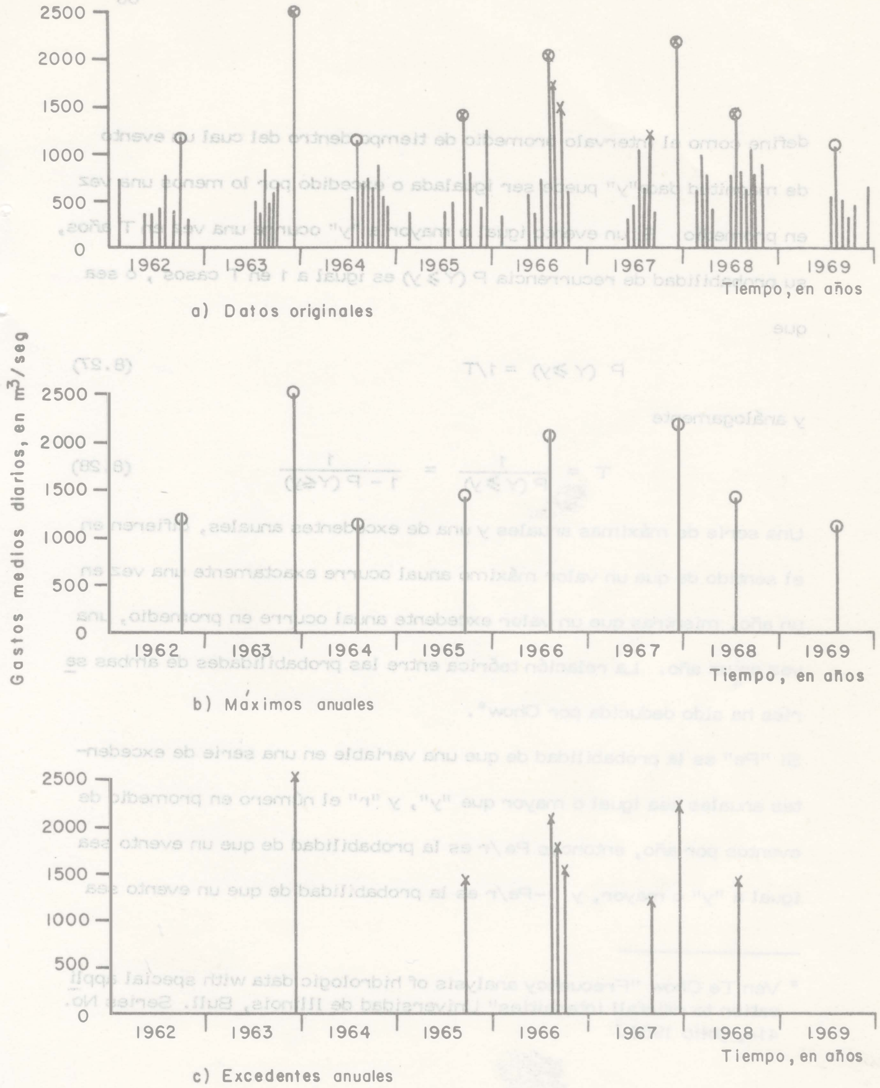


Fig. 8.7 Arreglo de datos hidrológicos en orden de ocurrencia

define como el intervalo promedio de tiempo dentro del cual un evento de magnitud dada "y" puede ser igualada o excedido por lo menos una vez en promedio. Si un evento igual o mayor a "y" ocurre una vez en T años, su probabilidad de recurrencia $P(Y \geqslant y)$ es igual a 1 en T casos, o sea que

$$P(Y \geqslant y) = 1/T \tag{8.27}$$

y análogamente

$$T = \frac{1}{P(Y \geqslant y)} = \frac{1}{1 - P(Y \leqslant y)} \tag{8.28}$$

Una serie de máximas anuales y una de excedentes anuales, difieren en el sentido de que un valor máximo anual ocurre exactamente una vez en un año, mientras que un valor excedente anual ocurre en promedio, una vez en un año. La relación teórica entre las probabilidades de ambas series ha sido deducida por Chow*.

Si "Pe" es la probabilidad de que una variable en una serie de excedentes anuales sea igual o mayor que "y", y "r" el número en promedio de eventos por año, entonces Pe/r es la probabilidad de que un evento sea igual a "y" o mayor, y 1-Pe/r es la probabilidad de que un evento sea

---

* Ven Te Chow "Frecuency analysis of hidrologic data with special application to rainfall intensities" Universidad de Illinois, Bull. Series No. 414, julio 1953.

menor que "y". Así, la probabilidad de que un evento de magnitud "y", llegue a ser un máximo de los "r" eventos en un año es $(1 - Pe/r)^r$. Se demuestra que esta probabilidad, tiende a e^{-Pe} cuando "Pe" es - pequeña en comparación con "r", lo cual es cierto en la mayoría de los casos.

Por lo anterior, la probabilidad "Pm" de que una máxima anual sea - igualada o excedida es igual a

$$Pm = 1 - e^{-Pe} \tag{8.29}$$

Si Tm y Te son respectivamente los períodos de retorno para los valores máximos anuales y excedentes anuales, de la ec. 8.27 se tiene que $Pm = {}^1/Tm$ y $Pe = {}^1/Te$, los que sustituyendo en 8.29 y ordenando, se tiene que

$$-\frac{1}{Te} = \log_e \left(\frac{Tm-1}{Tm}\right) \tag{8.30}$$

En esta ecuación se ve que para los valores de T mayores de 10 años, los valores de Tm y Te son practicamente iguales, por lo que para propósitos prácticos se puede concluir que si se dispone de un registro mayor de 10 años, el análisis de datos hidrológicos no difiere sensiblemente ya sea que se use para ello una serie de máximas anuales o bien una serie de excedentes anuales.

8.4.3 Períodos de retorno de los datos

Para el análisis de datos hidrológicos en relación con sus períodos de retorno, se requiere por una parte conocer la probabilidad de recurrencia de la distribución observada y por otra, la probabilidad de recurrencia correspondiente al evento que se desee obtener, ligado a la distribución teórica de mejor ajuste.

La obtención de los períodos de retorno ligados a los datos, se pueden valuar a partir de diversos criterios, pocos de los cuales tienen una explicación teórica. A continuación se analiza la obtención de las expresiones más usuales para valuar los períodos de retorno, refiriéndose estos a series de máximas y excedentes anuales.

Sea que se dispongan de n observaciones dato de una cierta distribución de eventos. Estos n valores se pueden arreglar en orden de magnitudes diferentes, asignándole a cada uno un número de orden m, el cual para el valor más grande es igual a uno, para el siguiente es dos, etc. Puede demostrarse * que la media $\bar{x}$ del número de veces que el m-avo valor más grande puede ser igualado o excedido en N futuros tanteos es

$$\bar{x} = N \frac{m}{n+1} \tag{8.31}$$

* Ven Te Chow "Frequency analysis of hidrologic date with special application torainfallintesities" Universidad de Illinois, Bull Series No.414, Appendix 1, Julio de 1953.

Para valores máximos anuales, Tm puede definirse como el tiempo en años para N futuros tanteos de que el m-avo valor más grande de los máximos anuales puede ser igualado o excedido una vez en promedio. Lo anterior implica que Tm = N cuando $\bar{x} = 1$. Sustituyendo estos valores en la ec 8.31 se tiene que

$$Tm = \frac{n+1}{m} \tag{8.32}$$

Esto indica que el período de retorno de un valor máximo anual es igual al número de años de registro más uno, dividido entre el número de orden. En el caso de los valores excedentes anuales, como se refieren a los valores más grandes, N/(n + 1) tiende a N/n y la ec 8.32 se tranforma en

$$T_e = \frac{n}{m} \tag{8.33}$$

que implica que el período de retorno de un valor excedente anual es igual al número de años de registro dividido entre el número de orden. Utilizando la ec 8.32 o la ec 8.33, ya sea que se analicen valores máximos o excedentes anuales, es posible disponer de una relación entre los valores en estudio y sus períodos de retorno o bien su frecuencia de incidencia. Generalmente, el siguiente paso en el análisis de datos hidrológicos es conocer la distribución de dichos valores y posteriormente inferir, basándose en una cierta frecuencia de incidencia, el valor que se puede presentar.

Antes de proceder al estudio del ajuste de los datos por medio de una cierta distribución, conviene ver la otra concepción del período de retorno, la cual está ligada al evento que en última instancia se desea obtener del análisis de estos datos para el diseño de una obra.

Ejemplo 8.11 En la tabla 8.17 se muestran los gastos máximos anuales en el Río Fuerte aforados por la Estación Hidrométrica Las Cañas, situada 4 km aguas abajo de la Presa Miguel Hidalgo en el Edo. de Sinaloa y correspondiente a la cuenca hidrológica del Río Fuerte. Se desea obtener para cada uno de los gastos su correspondiente período de retorno en términos de máximos anuales y de excedentes anuales.

Para el cálculo de los períodos de retorno "T" asociados a la serie de máximos anuales (Tm) y a la serie de duración parcial o excedentes anuales (Te) se elaboró la Tabla 8.18

Los valores de "Tm" y "Te" se obtuvieron empleando las ecuaciones (8.32) y (8.33) respectivamente, con $n = 18$ años de registro.

8.4.4 Períodos de retornos de los eventos de diseño

Si Pm es la probabilidad de que un valor máximo anual "y" sea igualado o excedido en un año, (1-Pm) es la probabilidad de que no se sobrepase "y" en un año en particular. Análogamente, $(1-Pm)^N$ es la probabilidad de que "y" sea igualado o excedido en N años, y por lo tanto

TABLA 8.17 Gastos Máximos Anuales en orden de ocurrencia en el Río Fuerte, Sinaloa. Valores en (m^3/seg).

ORDEN \ AÑO	1952	1953	1954	1955	1956	1957	1958	1959	1960
1	457.0	111.0	766.0	7477.0	659.0	110.9	161.7	169.0	8562.0
2	2264.0	1079.0	822.0	400.0	107.0	110.0	3888.0	1624.0	235.6
3	2229.0	1210.0	1619.0	1124.6	114.0	113.0	800.0	1127.6	212.4
4	324.4	857.6	1167.6	544.9	123.7	120.0	190.0	1202.9	268.8
5	332.0	53.8	651.6	694.0	161.0	129.0	303.1	343.5	709.4
MAXIMO ANUAL	2264.0	1210.0	1619.0	7477.0	659.0	129.0	3888.0	1624.0	8562.0
MAXIMOS EXCEDENTES ANUALES	2264.0 2229.0	1210.0 1079.0 857.6	1619.0 1167.6	7477.0 1124.6	———	———	3888.0	1624.0 1202.9 1127.6	8562.0

ORDEN \ AÑO	1961	1962	1963	1964	1965	1966	1967	1968	1969
1	452.0	166.8	259.8	469.3	171.6	192.1	188.1	203.7	213.0
2	298.2	187.0	477.0	224.9	147.0	206.8	208.1	225.0	207.9
3	697.2	187.0	460.9	436.4	152.0	1507.0	174.0	740.0	175.3
4	511.0	210.0	456.3	607.4	169.2	1251.8	168.8	1000.0	168.5
5	594.6	328.8	824.0	222.0	169.2	439.2	595.0	685.0	175.3
MAXIMO ANUAL	697.2	328.8	824.0	607.4	171.6	1507.0	595.0	1000.0	213.0
MAXIMOS EXCEDENTES ANUALES	———	———	824.0	———	———	1507.0 1251.8	———	1000.0	———

Tabla 8.18 Ordenamiento de cálculo.

Año	Gasto Máximo Anual de Mayor a Menor	Número de orden m	Tm (años)	Año	Gastos Máximos Excedentes Anuales de Mayor a Menor	Número de orden m	Te (años)
1960	8562.0	1	19.00	1960	8562.0	1	18.00
1955	7477.0	2	9.50	1955	7477.0	2	9.00
1958	3888.0	3	6.33	1958	3888.0	3	6.00
1952	2264.0	4	4.75	1952	2264.0	4	4.50
1959	1624.0	5	3.80	1952	2229.0	5	3.60
1954	1619.0	6	3.17	1959	1624.0	6	3.00
1966	1507.0	7	2.71	1954	1619.0	7	2.57
1953	1210.0	8	2.38	1966	1507.0	8	2.25
1968	1000.0	9	2.11	1966	1251.8	9	2.00
1963	824.0	10	1.90	1953	1210.0	10	1.80
1961	697.2	11	1.73	1959	1202.9	11	1.64
1956	659.0	12	1.58	1954	1167.6	12	1.50
1964	607.4	13	1.46	1959	1127.6	13	1.38
1967	595.0	14	1.36	1955	1124.6	14	1.29
1962	328.8	15	1.27	1953	1079.0	15	1.20
1969	213.0	16	1.19	1968	1000.0	16	1.13
1965	171.6	17	1.12	1953	857.6	17	1.06
1957	129.0	18	1.06	1963	824.0	18	1.00

$$Pn = 1 - (1 - Pm)^N$$

De la ec 8.27, Pm = 1/Tm, por lo que

$$Pn = 1 - (1 - 1/Tm)^N = 1 - (1 - N/Tm)$$

en donde

$$Tm = \frac{N}{Pn} \tag{8.34}$$

que implica que el período de retorno ligado al evento que se quiera obtener, es función del intervalo de tiempo en el cual se desea que no sea igualado dicho evento entre la probabilidad de que si sea igualado o superado. Así por ejemplo, para el diseño del vertedor de una presa, si se tiene que a dicha obra se le asigna una vida útil de 50 años, y si se considera que la avenida de diseño del vertedor tenga una probabilidad de 0.01 de que se presente o sea superada durante ese lapso, la avenida de diseño tendrá, de acuerdo con la ecuación anterior un período de retorno de 5,000 años.

En el caso de trabajar con series de duración parcial y valores de excedentes anuales, el período de retorno para determinar el evento de diseño tiene una expresión similar a la ec (8.34).

Si se considera que la vida útil de una cierta obra es constante, de la ec (8.34) se tiene que el período de retorno es función inversa de la pro babilidad de ocurrencia del evento. Para valuar la probabilidad de ocurrencia, lo que se podría llamar también la probabilidad de riesgo, se requiere tener en cuenta: (a) Costo de la obra; (b) Daños que se pueden tener al presentarse una falla; (c) Costo de Mantenimiento; (d) Inconvenientes y perjuicios en caso de que falle la obra; (e) riesgo de vidas humanas.

De lo anterior se ve el porqué de la diferencia entre los períodos de retorno de obras de drenaje y las presas. En general si falla una obra de drenaje los percances ocasionados son mínimos comparados con los originados por la falla de una presa. Así, en el caso de diseño de vertedores de presas los períodos de retorno son muy grandes; en cambio, en el caso de puentes importantes fluctúa usualmente de 100 a 200 años, en el caso de obras de alcantarillado para drenaje de caminos de 25 a 50 años, drenaje en ciudades y aeropuertos de 5 a 10 años.

8.5 Distribuciones de datos hidrológicos

Conocidos los períodos de retorno correspondientes a cada uno de los datos de una muestra, es posible así proceder a la obtención de su distribución de probabilidades y hacer inferencias.

En la Tabla 8.1 se indican las principales distribuciones teóricas de probabilidades utilizadas en hidrología, así como sus características.

Por otra parte en el subinciso 8.1.4 se vió como se puede conocer e interpretar si alguna distribución teórica seleccionada es representativa de los datos de una muestra. En este inciso se plantean las distribuciones más usuales y su relación entre las características de un cierto evento hidrológico y sus frecuencias ó períodos de retorno.

8.5.1 Distribución de valores extremos Tipo I

Esta distribución fué propuesta por Gumbel* para el análisis de frecuencias de avenidas. Considerando que cada máximo anual es el valor extremo observado en una muestra de un año, se tendrán si se dispone de un número infinito de muestras anuales, que la probabilidad acumulada P (y) de cualquiera de los extremos (máximos anuales) sea menor a la variable ilimitada "y", se aproxima a la expresión.

$$P\ (y) = e^{-e^{-(a+y)/c}} \qquad (8.35)$$

donde "a" y "c" son parámetros estadísticos, cuyos valores obtenidos por el método de los momentos, para una población infinita, se calculan como:

$$a = 0.5772\ c - \bar{y} \qquad (8.36)$$

$$c = (\sqrt{6}/\pi)\ Sy$$

donde $\bar{y}$ es la media y Sy la desviación estandar de los valores de la población.

Como la muestra es siempre finita, Gumbel considera que

---

* E. J. Gumbel "The Return Period of Floods Flows", Annals Mathematical Statistics, Vol XII, No. 2. junio 1941.

$$a = \bar{y}_N\, c - \bar{y} \qquad (8.36)$$

y:

$$c = Sy / \sigma_N \qquad (8.37)$$

donde $\bar{y}_N$ y σ_N son valores solo función del tamaño de la muestra. Sus valores se indican en la Tabla 8.19

La probabilidad complementaria de P (y) es la probabilidad Pm de que un máximo anual de magnitud "y" sea igualado ó excedido. Por lo tanto:

$$Pm = 1 - P(y) = 1 - e^{-e^{-(a+y)/c}}$$

y teniendo en cuenta la ec 8.27, la ecuación anterior se puede expresar como

$$\frac{Tm - 1}{Tm} = e^{-e^{-(a+y)/c}}$$

trasponiendo y simplificando, se llega a que:

$$Y = -a - c \log_e \log_e \frac{Tm}{Tm - 1} \qquad (8.38)$$

Sustituyendo en esta ecuación las ec 8.36 y 8.37, y ordenando se tiene que

$$y = \bar{y} - \frac{Sy}{\sigma_N} (\bar{y}_N + \log_e \log_e \frac{Tm}{Tm - 1}) \qquad (8.39)$$

expresión que permite hacer inferencias de los valores de y para cualquier período de retorno Tm, conociendo el tamaño N de la muestra de datos de máximos anuales, su media $\bar{y}$ y su desviación estándar Sy.

Tabla 8.19 Valores de $\overline{Y}_N$ y σ_N

N	$\overline{Y}_N$	σ_N	N	$\overline{Y}_N$	σ_N
8	.4843	.9043	49	.5481	1.1590
9	.4902	.9288	50	.54854	1.16066
10	.4952	.9497	51	.5489	1.1623
11	.4996	.9676	52	.5493	1.1638
12	.5035	.9833	53	.5497	1.1653
13	.5070	.9972	54	.5501	1.1667
14	.5100	1.0095	55	.5504	1.1681
15	.5128	1.02057	56	.5508	1.1696
16	.5157	1.0316	57	.5511	1.1708
17	.5181	1.0411	58	.5515	1.1721
18	.5202	1.0493	59	.5518	1.1734
19	.5220	1.0566	60	.55208	1.17467
20	.52355	1.06283	62	.5527	1.1770
21	.5252	1.0696	64	.5533	1.1793
22	.5268	1.0754	66	.5538	1.1814
23	.5283	1.0811	68	.5543	1.1834
24	.5296	1.0864	70	.55477	1.18536
25	.53086	1.09145	72	.5552	1.1873
26	.5320	1.0961	74	.5557	1.1890
27	.5332	1.1004	76	.5561	1.1906
28	.5343	1.1047	78	.5565	1.1923
29	.5353	1.1086	80	.55688	1.19382
30	.53622	1.11238	82	.5572	1.1953
31	.5371	1.1159	84	.5576	1.1967
32	.5380	1.1193	86	.5580	1.1980
33	.5388	1.1226	88	.5583	1.1994
34	.5396	1.1255	90	.55860	1.20073
35	.54034	1.12847	92	.5589	1.2020
36	.5410	1.1313	94	.5592	1.2032
37	.5418	1.1339	96	.5595	1.2044
38	.5424	1.1363	98	.5598	1.2055
39	.5430	1.1388	100	.56002	1.20649
40	.54362	1.14132	150	.56461	1.22534
41	.5442	1.1436	200	.56715	1.23598
42	.5448	1.1458	250	.56878	1.24292
43	.5453	1.1480	300	.56993	1.24786
44	.5458	1.1499	400	.57144	1.25450
45	.54630	1.15185	500	.57240	1.25880
46	.5468	1.1538	750	.57377	1.26506
47	.5473	1.1557	1000	.57450	1.26851
48	.5477	1.1574	∞	.57722	1.28255

La ec 8.39 se puede transformar en

$$y = \bar{y} + S_y K \tag{8.40}$$

donde

$$K = -\frac{1}{\sigma_N}\left(\bar{Y}_N + \log_e \log_e \frac{T_m}{T_m - 1}\right) \tag{8.41}$$

denominado por Chow* como factor de frecuencia, el cual depende del tamaño de la muestra de datos y del período que se está analizando. - En la Tabla 8.20 proporcionar algunos valores de K

En el caso de utilizar excedentes anuales en lugar de máximos anuales, se pueden utilizar las expresiones anteriores solo teniendo en cuenta la ec 8.30. Así, para excedentes anuales la ec 8.39 se transforma en

$$Y = \bar{Y} - \frac{S_y}{\sigma_N}\left(\bar{Y}_N - \log_e T_e\right) \tag{8.42}$$

Observese que tanto la ec(8.39)como la (8.42), si se grafican en un papel de probabilidades especial, y contra Tm o Te, se obtiene la ecuación de una recta. El hecho de que estas ecuaciones representan una línea - recta, no implica que los datos de la muestra que se está analizando - esten sobre la línea, por lo que es necesario conocer el intervalo de confianza de los resultados obtenidos del análisis de frecuencias. Así para cierto valor Tm, se tendrá que

$$y - \Delta y < y < y + \Delta y$$

*Ven Te Chow " A General Formula for Hydrologic Frequency Analysis". Trnas. Amer. Geophys. Union. Vol 32, No.2, abril 1952.

Tabla 8.20 Valor del factor de frecuencia K para diferentes tamaños de muestra y período de retorno.

T años / n años	5	10	15	20	25	30	50	60	75	100
15	0.967	1.703	2.117	2.410	2.632	2.823	3.321	3.501	3.721	4.005
20	0.919	1.625	2.023	2.302	2.517	2.690	3.179	3.352	3.563	3.836
25	0.888	1.575	1.963	2.235	2.444	2.614	3.088	3.257	3.463	3.729
30	0.866	1.541	1.922	2.188	2.393	2.560	3.026	3.191	3.393	3.653
35	0.851	1.516	1.891	2.152	2.354	2.520	2.979	3.142	3.341	3.598
40	0.838	1.495	1.866	2.126	2.326	2.489	2.943	3.104	3.301	3.554
45	0.829	1.478	1.847	2.104	2.303	2.464	2.913	3.078	3.268	3.520
50	0.820	1.466	1.831	2.086	2.283	2.443	2.889	3.048	3.241	3.491
55	0.813	1.455	1.818	2.071	2.267	2.426	2.869	3.027	3.219	3.467
60	0.807	1.446	1.806	2.059	2.253	2.411	2.852	3.008	3.200	3.446
65	0.801	1.437	1.796	2.048	2.241	2.398	2.837	2.992	3.183	3.429
70	0.797	1.430	1.788	2.038	2.230	2.387	2.824	2.979	3.169	3.413
75	0.792	1.423	1.780	2.029	2.220	2.377	2.812	2.967	3.155	3.400
80	0.788	1.417	1.773	2.020	2.212	2.368	2.802	2.956	3.145	3.387
85	0.785	1.413	1.767	2.013	2.205	2.361	2.793	2.946	3.135	3.376
90	0.782	1.409	1.762	2.007	2.198	2.353	2.785	2.938	3.125	3.367
95	0.780	1.405	1.757	2.002	2.193	2.347	2.777	2.930	3.116	3.357
100	0.779	1.401	1.752	1.998	2.187	2.341	2.770	2.922	3.109	3.349

donde y es el valor obtenido de la ec (8.39) y Δy el intervalo de confianza con una cierta probabilidad.

Gumbel propone para calcular los intervalos de confianza con una probabilidad del 68%, lo siguiente:

1) Para el valor más grande de la muestra analizada (número de orden m = 1)

$$\Delta y_1 = S_y F(N) \tag{8.42a}$$

donde Sy es la desviación estándar de la muestra y F (N) es función del tamaño de la muestra N (fig 8.8.a)

2) Para el segundo valor más grande (número de orden m = 2)

$$\Delta y_2 = \frac{0.661\,(N+1)}{N-1}\,\Delta y_1 \tag{8.42b}$$

3) Para los otros valores de la muestra

$$\Delta y = \frac{0.877}{N}\,\Delta y_1\,F(Tm) \tag{8.42c}$$

donde F (Tm) es una función de Tm, cuyos valores si Tm ≤ 10 años se encuentran en la fig 8.8.b. Para valores de Tm mayores de 10 años, se tiene que:

$$F(Tm) = Tm^{0.5} \tag{8.42d}$$

4) Para extrapolar a valores mayores al máximo de la muestra, el intervalo de confianza se considera constante e igual a Δy_1

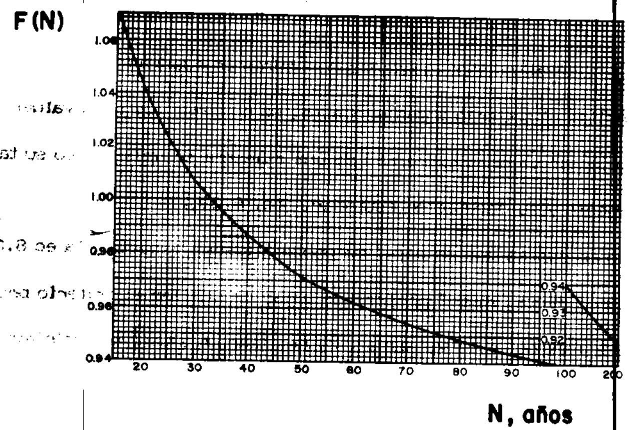


Fig.8.8a Relación entre N y F(N)

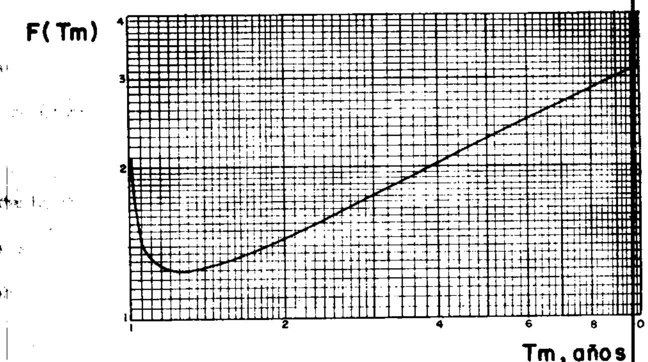


Fig.8.8b Relación entre Tm y F(Tm)

(M) ꟻ

8.5.2 Distribuciones ajustadas por mínimos cuadrados

Si se analiza el criterio antes expuesto, se observa que para valuar la distribución de valores extremos de una muestra se utiliza solo su tamaño, su media y su desviación estándar. (ec. 8.39).

Teniendo en cuenta que en papel de probabilidades especial, la ec 8.39 es una línea recta, Nash* propone en lugar de utilizar el criterio propuesto por Gumbel para valuar a y c, realizar un ajuste por mínimos cuadrados.

La ec 8.38 se puede transformar en

$$y - a - c \log_e \log_e \frac{Tm}{Tm-1} = e + f x \qquad (8.43)$$

en donde los parámetros e y f se determinan de acuerdo con las ec (8.15) y (8.16) y el intervalo de confianza según la ec (8.20)

De la ecuación 8.24 se ve que también es factible considerar que:

$$y = g + h \log T_e = g + hx \qquad (8.44)$$

expresión análoga a la anterior, solo que ahora el ajuste es una línea recta en papel semilogarítmico.

De esta manera es posible escoger diferentes tipos de curvas de ajuste de los datos y seleccionar la mejor. En este caso, si la correlación es simple, se aceptará como mejor la que tenga el máximo coeficiente de correlación.

---

* R.B. Thorn, "River Engineering and Water Conservation Works", -- Butter worths (1966)

Ejemplo 8.12 A partir de la información del ejemplo 8.11 deducir el gasto máximo que se puede presentar en la Estación Hidrométrica Las Cañas sobre el Río Fuerte, considerando un período de retorno de 100 año.

Según la ec. (8.39) y (8.42) Gumbel establece la igualdad de:

$$y = \bar{y} - \frac{Sy}{\sigma N}\left[\bar{y}_N + \log_e \log_e \frac{Tm}{Tm-1}\right] = f \text{ (serie de máximos anuales)}$$

$$y = \bar{y} - \frac{Sy}{\sigma N}\left[\bar{y}_N - \log_e Te\right] = f \text{ (excedentes anuales)}$$

donde

y = Qmáx esperado para un cierto período de retorno Tr

Para valuar las incógnitas de las ecuaciones anteriores, considerando una serie de máximos anuales, en la Tabla 8.21 se muestra el ordenamiento del cálculo.

Así, la media de los gastos, ec (8.6), $\bar{y} = \frac{33376.0}{18} = 1854.22$ m^3/seg.

su desviación estándar, ec (8.8), $Sy = \sqrt{\frac{100084648.58}{18-1}} = \sqrt{5887332.26}$

$Sy = 2426.39$

De la Tabla 8.19 se obtiene para N = 18 años, los parámetros $\bar{y}_N$ y σ_N siendo estos:

T A B L A 8.21

ORDENAMIENTO DE CALCULOS PARA LA OBTENCION DE $\bar{y}$ (MEDIA) Y Sy (DESV. ESTANDAR DE LA MUESTRA

Orden	Año de Observación	Gasto Máximo Anual Aforado yi (m^3/s)	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	1952	2264.00	409.78	167919.64
2	1953	1210.00	-644.22	415019.40
3	1954	1619.00	-235.22	55328.44
4	1955	7477.00	5622.78	31615654.92
5	1956	659.00	-1195.22	1428550.84
6	1957	129.00	-1725.22	2976384.04
7	1958	3888.00	2033.78	4136261.08
8	1959	1624.00	-230.22	53001.24
9	1960	8562.00	6707.78	44994312.52
10	1961	697.20	-1157.02	1338695.28
11	1962	328.80	-1525.42	2326906.17
12	1963	824.00	-1030.22	1061353.24
13	1964	607.40	-1246.82	1554560.11
14	1965	171.60	-1682.62	2831210.06
15	1966	1507.00	-347.22	120561.72
16	1967	595.00	-1259.22	1585635.00
17	1968	1000.00	-854.22	729691.80
18	1969	213.00	-1641.22	2693603.08
n 18	S U M A	33376.00		100084648.58

$$\bar{y}_N = 0.5202$$

$$\sigma_N = 1.0493$$

Finalmente, para Tm = 100 años, se deduce que

$$y = 1854.22 - \frac{2426.39}{1.0493}\left(0.5202 + \log_e \log_e \frac{100}{100-1}\right)$$

$$y = 1854.22 - 2312.69\ (0.5202 + (-4.600))$$

$$y = 1854.22 + 9434.09 = 11\,288.31$$

De esta manera, el gasto máximo obtenido resulta

$$y = Qmax = 11{,}288.31\ m^3/seg$$

$$Tm = 100$$

Para calcular el intervalo de confianza y, de la fig (8.8.b) se tiene - para N = 18 años que F (N) = 1.056

Substituyendo en la ec 8.42 a se deduce

$$y_1 = 2426.39\ (1.056) = 2562.27\ m^3/seg = Q$$

De esta manera, el gasto máximo de diseño resultaría:

$$Q_{diseño} = Qmáx \pm Q = 11288.31 \pm 2562.27$$

lo cual, considerando la condición más desfavorable resulta de 13,850 m^3/seg.

Aplicando el criterio de Nash, de la ec (8.43) se tiene

$$y = -a + c \log_e \log_e \frac{Tm}{Tm-1} = e + f\,x$$

$$y = Qmáx$$

$$e = -a;\ f'\ (Qmáx, Tm)$$

$$f = c;\ f''\ (Qmáx, Tm)$$

$$x = \log_e \log_e \frac{Tm}{Tm-1}$$

por lo que solo se necesita calcular el valor de los parámetros e y f (a y c), tales que minimizen el error entre los puntos de la muestra y la recta de regresión, ec (8.14). De acuerdo a las ecuaciones (8.15) y (8.16)

$$f = \frac{Sxy}{Sxx}$$

$$e = \bar{y} - f\bar{x}$$

En la Tabla 8.22 se muestra el ordenamiento de los cálculos. Así, de la ec (8.17),

$$Sxx = 18\ (24.5558) - (-9.4037)^2 = 442.0044 - 88.4295$$

$$= 353.5749$$

$$Syy = 18\ (161971169.60) - (33376.0)^2$$

$$= 2915471052.8 - 1113957376 = 1801513676.80$$

y de la ec (8.18)

$$Sxy = 18\ (-56891.87) - (-9.4037)\ (33376.0)$$

$$= -1024053.66 + 313857.8912 = -710195.7688$$

Substituyendo se deduce que

$$f = \frac{-710195.77}{353.57} = -2006.20$$

T A B L A 8.22

ORDENAMIENTO DE VALORES PARA EL CALCULO DE LOS PARAMETROS DE LA RECTA DE REGRESION Y COEFICIENTE DE CORRELACION

Orden	Qmax anual Y_i-(m^3/s)	Tm (años)	$\frac{Tm}{Tm-1}$	$X_i = \log_e \log_e \frac{Tm}{Tm-1}$	$(X_i)^2$	$(Y_i)^2$	$(X_i)(Y_i)$
1	8562.00	19.00	1.0556	- 2.9175	8.5120	73307844.00	- 24979.87
2	7477.00	9.50	1.1176	- 2.1962	4.8233	55905529.00	- 16420.95
3	3888.00	6.33	1.1876	- 1.7606	3.0996	15116544.00	- 6845.05
4	2264.00	4.75	1.2667	- 1.4423	2.0802	5125696.00	- 3625.32
5	1624.00	3.80	1.3571	- 1.1862	1.4071	2637376.00	- 1926.38
6	1619.00	3.17	1.4608	- 0.9702	0.9413	2621161.00	- 1570.77
7	1507.00	2.71	1.5848	- 0.7755	0.6015	2271049.00	- 1168.74
8	1210.00	2.38	1.7246	- 0.6069	0.3684	1464100.00	- 734.40
9	1000.00	2.11	1.9009	- 0.4427	0.1959	1000000.00	- 442.66
10	824.00	1.90	2.1111	- 0.2914	0.0849	678976.00	- 240.12
11	697.20	1.73	2.3699	- 0.1475	0.0218	486087.84	- 102.86
12	659.00	1.58	2.7241	0.0021	0.0000	434281.00	1.42
13	607.40	1.46	3.1739	0.1441	0.0208	368934.76	87.51
14	595.00	1.36	3.7778	0.2845	0.0810	354025.00	169.29
15	328.80	1.27	4.7037	0.4372	0.1911	108109.44	143.75
16	213.00	1.19	6.2632	0.6069	0.3683	45369.00	129.26
17	171.60	1.12	9.3333	0.8036	0.6458	29446.56	137.90
18	129.00	1.06	17.6667	1.0549	1.1128	16641.00	136.08
n=18	Y_i =33376.00	S U M A		-9.4037	24.5558	161971169.60	- 56891.87

y como

$$\bar{x} = \frac{-9.4037}{18} = -0.5224; \quad \bar{y} = \frac{33376}{18} = 1854.22$$

se determina

$$e = 1854.22 - (-2006.2)\ (-0.5224) = 1854.22 - 1048.04 = 806.18$$

Con esta información la ec. de la recta de regresión resultante es

$$y' = e + fx$$

$$y' = 806.18 - 2006.20\ x$$

y como $e = -a; \quad a = -806.18$

$$f = c; \quad c = -2006.20$$

la ecuación buscada (8.43) se expresa

$$y' = 806.18 - 2006.20 \log_e \log_e \frac{Tm}{Tm - 1}$$

con lo cual

$$Qmáx = 806.18 - 2006.20 \log_e \log_e \frac{Tm}{Tm - 1}$$

El coeficiente de correlación asociado a los puntos dato y recta de regresión, resulta según la ec (8.19) con un valor de

$$r = \frac{-710195.77}{(353.57)\ (1801513676.80)^{1/2}} = \frac{-710195.77}{798098.48}$$

$$= -0.8899 = -0.9$$

Con la ecuación anterior, para Tm = 100 años se obtiene un

$$Qmáx = 10034.70\ m^3/seg.$$

y el intervalo de confianza asociado, si se elige un nivel de significancia de 0.95 $\alpha = 0.05; \frac{\alpha}{2} = 0.025$ y grados de libertad $\nu = n-2 = 18-2 = 16$, de los valores tabulados para la distribución "t" se obtiene:

$$t\,\alpha/2 = t_{0.025} = 2.12$$

y la variancia del error, ec (8.19)

$$Se^2 = \frac{1801513676.8}{18\,(18-2)} \left[1 - \frac{(-710195.77)^2}{(353.57)(1801513676.8)} \right] = 6255255.82\,(0.2082) = 1302097.09$$

$$Se = \sqrt{Se^2} = 1141.095$$

Substituyendo en la ec(8.20)

$$\mathcal{E} = \pm\, 2.12 \quad (1141.095) \sqrt{1 + \frac{1}{18} + \frac{18\,[-4.6-(-0.5224)]^2}{353.5749}}$$

Por lo que

$$\mathcal{E} = \pm\, 2419.12 \sqrt{1 + 0.0556 - 0.8465} = \pm\, 2419.12 \sqrt{0.4573}$$

$$= \pm\, 2419.12 \quad (0.6762) \quad = \pm\, 1635.90$$

y finalmente el gasto máximo de diseño más probable sería

$$Qmáx = 10034.70 + 1635.90 \qquad Q_D = 11671 \text{ m}^3/\text{seg}$$

$$Tm = 100 \text{ años}$$

condición más desfavorable.

En la figura 8.9 se muestra la relación existente entre la información disponible y los ajustes, tanto considerando la distribución de valores extremos como el propuesto por Nash por mínimos cuadrados.

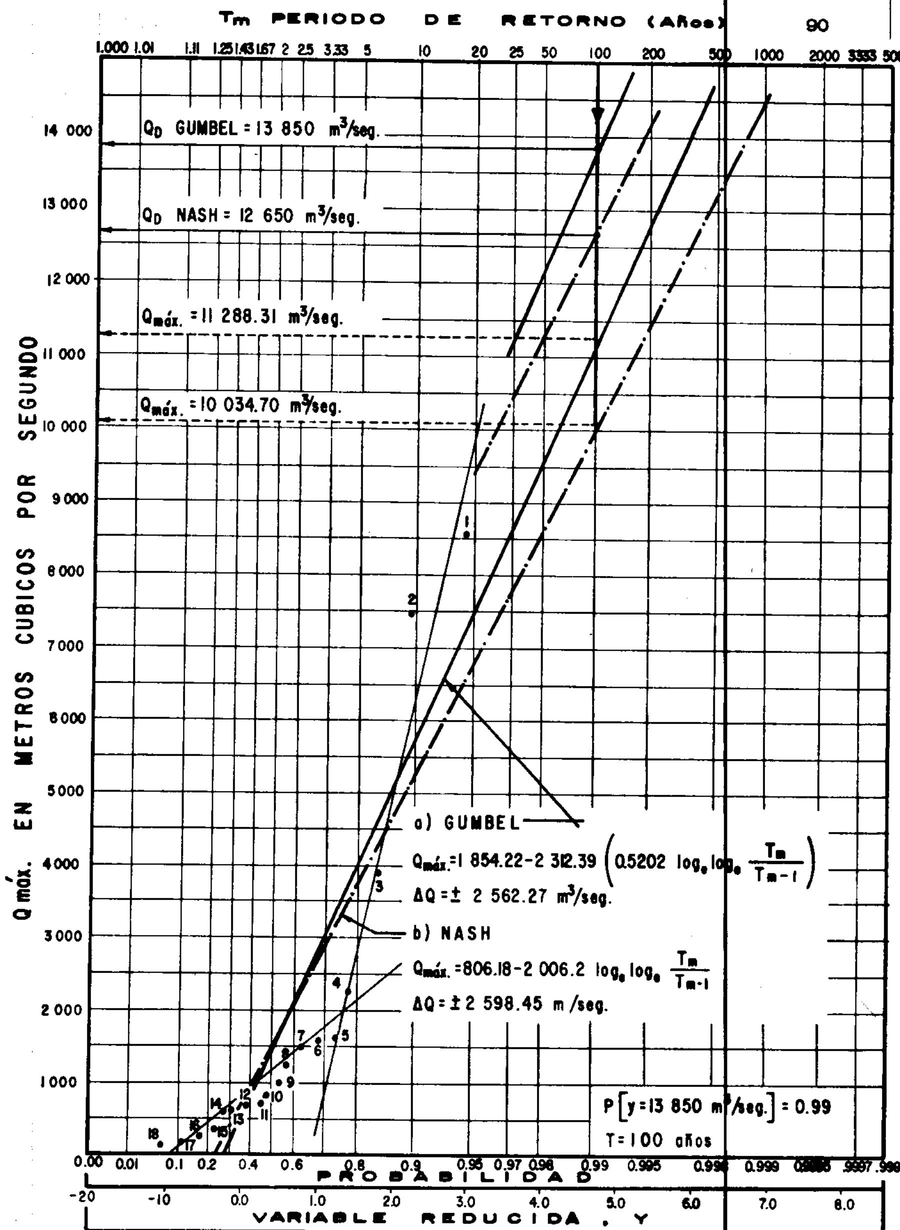


Fig. 8. 9 Rectas de ajuste utilizando las distribuciones a) Gumbel , b) Nash

8.5.3 Distribución de frecuencias aplicada a dos poblaciones

Muchas veces, cuando se analiza una muestra de datos hidrológicos, resulta que dicha muestra tiene dos poblaciones con características distintas. Por ejemplo, cuando se analizan gastos máximos anuales, en ocasiones se tiene que algunos ocurren durante la época de ciclones y otros no, lo que ocasiona lo antes indicado.

Por lo anterior, conviene antes de proceder a ajustar una cierta distribución de frecuencias a la información, conviene primero graficar los datos con respecto a sus períodos de retorno (fig 8.9). De esta manera, se puede apreciar si los datos son de una sola población o de dos.

En el caso de tener dos poblaciones, González,* aplicando por separado la distribución propuesta por Gumbel (ec 8.35) a cada una de las poblaciones, obtiene una función de distribución de probabilidades para gastos máximos anuales, del tipo

$$P(y) = e^{-e^{-\frac{y+a1}{c1}}}\left[p + (1-p)\, e^{-e^{-\frac{y+a2}{c2}}}\right] \qquad (8.45)$$

donde p, a_1, c_1, a_2, c_2, son parámetros por estimar. Para esto, utiliza el método matemático del máximo ascenso, tal que, los parámetros

---

* Fernando J. González V. "Contribución del análisis de frecuencias de valores extremos de los gastos máximos en un río". Publicación del Instituto de Ingeniería, UNAM Diciembre de 1970. (277).

produzcan una función teórica que haga mínima la suma de los cuadrados de las desviaciones entre los valores de esta función y los valores datos, es decir que minimizan el error.

Otra forma de proceder en el análisis de frecuencia de dos poblaciones es utilizar el criterio de mínimos cuadrados, previa división de ambas poblaciones, una vez graficados los valores de la muestra con respecto a sus períodos de retorno. Este criterio presenta el problema de los errores inherentes ocasionados al dividir ambas poblaciones, ya que muchas veces se tiene muy poca información para poder apreciar la división de las mismas.

Ejemplo 8.13 Se desea obtener el gasto máximo para el Río Fuerte, con los mismos datos del ejemplo 8.12, aplicando el criterio del ajuste de la distribución de valores extremos de dos poblaciones.

Analizando los datos disponibles fig(8.9), puede notarse apriori la existencia de las dos poblaciones definidas por este criterio; podría asegurarse evidentemente que el gasto máximo aforado de 8562.0 m^3/seg. corresponde a la población cuyos valores provienen de fenómenos ciclónicos, extraordinarios y el gasto de 1619.0 m^3/seg. pertenece a la población cuyos escurrimientos son función de las características atmosféricas meteorológicas reinantes de manera ordinaria o normal en la región;

aun más, siendo esta una zona costera podría afirmarse que la aplicación del método de las dos poblaciones resultaría el más adecuado. La ec (8.45) resulta una función implícita en "y" por lo que esta variable no es posible el despojarse y deberá de procederse para la obtención del gasto de manera indirecta, suponiendo un gasto y obteniendo su período de retorno o probabilidad correspondiente; de la misma forma dada la complejidad del aparato matemático para la correcta obtención optinizada de los parámetros propios se desarrollará solamente en este ejemplo una primera aproximación cualitativa en la determinación de los mismos considerando poblaciones infinitas para ambos casos.

Se definen de esta forma los parámetros (Sec. 8.5.1):

p = f (número de casos en que los gastos máximos graficados en papel de Gumbel se consideran de población no ciclónica)

Población no ciclónica:

$$a_1 = 0.577\ c_1 - \bar{x}_1$$
$$c_1 = (\sqrt{6/\pi})\ S_1$$

Población ciclónica

$$a_2 = 0.577\,c_2 - \bar{x}_2$$

$$c_2 = (\sqrt{6}/\pi)\,S_2$$

Donde $\bar{x}i$, Si, son la media y desviación estándar de los valores propios a cada población.

Cálculo de "p"

De la fig (8.9) tentativamente podrían separarse las dos rectas de cada población. Los gastos (en orden, tabla 8.22) de la población de tipo ciclónico serían entonces

ciclónico : 1, 2, 3 y 4

no ciclónico : del 5 al 18

$$p = \frac{14}{18} = 0.7778$$

Cálculo de ai , ci :

La media y desviación estándar correspondientes a cada población son (ecs (8.6) y (8.8)):

no ciclónica	ciclónica
1	1
$\bar{x}_1 = 798.929$	$\bar{x}_2 = 5547.750$
$S_1 = 524.737$	$S_2 = 2963.415$

y los parámetros

$$c_1 = (\sqrt{6}/\pi)S_1 = 409.136$$

$$a_1 = 0.577\,c_1 - \bar{x}_1 = -562.858$$

$$c_2 = (\sqrt{6}/\pi)S_2 = 2310.565$$

$$a_2 = 0.577\, c_2 - \bar{x}_2 = -4214.554$$

además el período de retorno asociado a nuestro problema fue de

$$T = 100 \text{ años}$$

$$P(Y) = 0.99$$

sustituyendo los parámetros en la ec (8.45) tanteando para encontrar el período necesario-requerido de manera indirecta se tiene:

1er. tanteo: Con el gasto obtenido por Gumbel en el ejemplo 8.12,

$$y = 11288.3 \text{ m}^3/\text{seg.}$$

$$P(Y) = e^{-e^{-\left(\frac{11288.3 - 562.858}{409.136}\right)}} \left\{0.7778 + 0.2222 \quad e^{-e^{-\left(\frac{11288.3-4214.554}{2310.565}\right)}}\right\}$$

$$= e^{-e^{-(26.215)}} \left[0.7778 + (0.2222)\, e^{-e^{-(3.0615)}}\right]$$

$$= e^{-0} \left[0.7778 + (0.2222)\, e^{-0.0468}\right]$$

$$= 0.7778 + 0.2120 = 0.9898$$

como $0.99 > 0.9898$

$T = 100 > 98$ años

2 do. tanteo: Con $y = 11{,}400 \text{ m}^3/\text{seg}$ análogamente

se obtiene $P(Y) = 0.9903$

3er. tanteo : Con $y = 11{,}300 \text{ m}^3/\text{seg}$

se obtiene $P(Y) = 0.99$

para $T = 100$ años

De esta forma, se concluye, que el gasto máximo para el Río Fuerte con un T = 100 años, aplicando el criterio de las dos poblaciones en una primera aproximación sería de:

$$Qmáx = 11,300 \ m^3/seg.$$

Conviene aplicar este método para el análisis y comparación de resultados obtenidos, cuando la importancia de la obra que se desea diseñar es considerable donde en caso de presas generalmente se asignen para diseño períodos de retorno de 1 000 a 10 000 años, lo que aunado a la aplicación de la matemática necesaria a la optimización de los parámetros de la distribución nos daría diferencias más notables entre ambos métodos.

En la fig (8.10) se muestra la relación gráfica de la distribución de frecuencia aplicada a dos poblaciones y obtenida para diferentes gastos máximos y su correspondiente probabilidad o período de retorno asociado.

8.6 Extrapolación de las curvas de frecuencias

Cuando la frecuencia o período de retorno de un evento hidrológico a determinar es considerablemente mayor que la longitud de registro disponible, se requiere extrapolar la distribución obtenida de los datos. Debido a la gran incertidumbre que involucra tales procedimientos, no se recomienda extrapolar para el diseño de grandes estructuras cuya falla

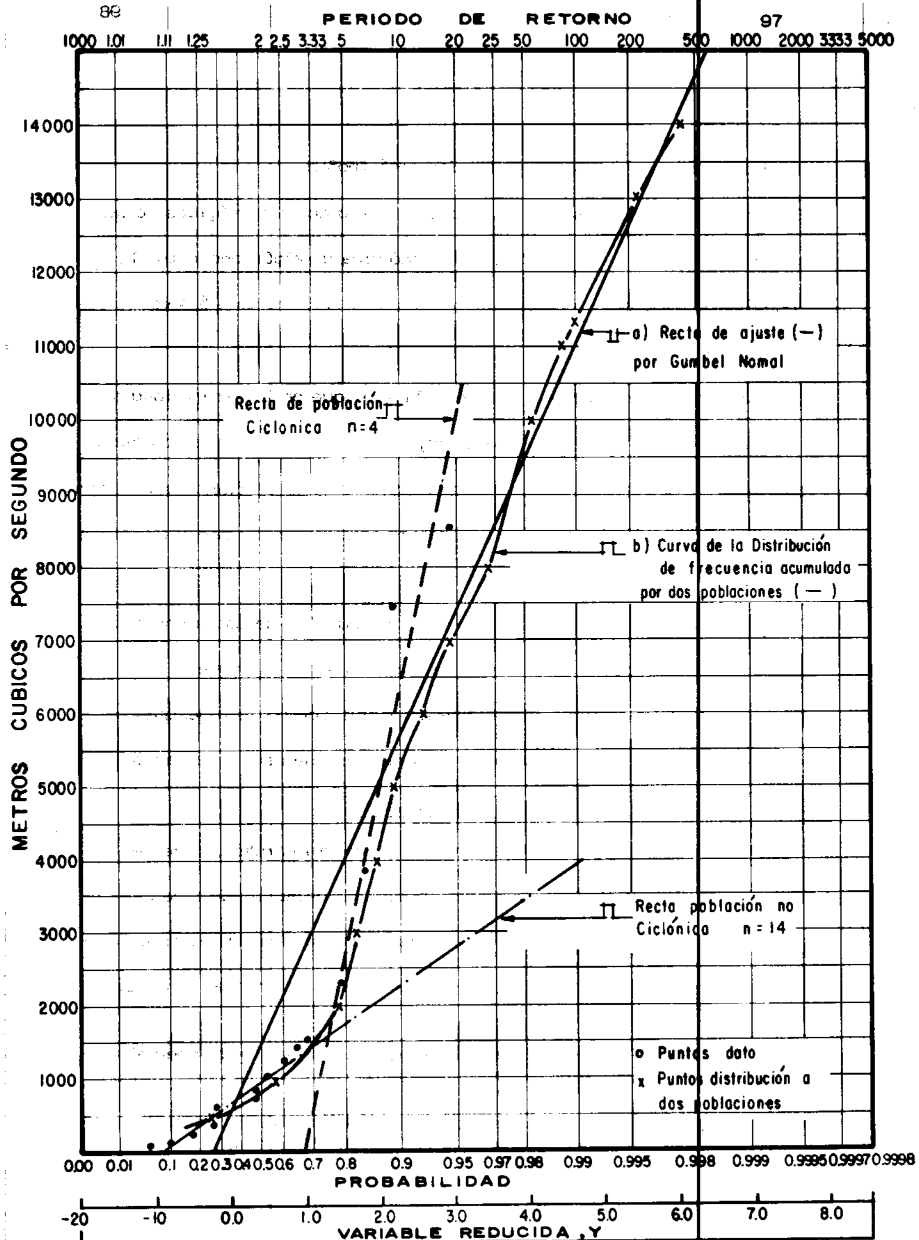


Fig.8.10 Comparación entre los valores reales y diferentes distribuciones ajustadas.

implique pérdidas humanas y grandes catástrofes.

Sin embargo, si el análisis de frecuencia de datos hidrológicos se utiliza para determinar la justificación económica de algún proyecto, la extrapolación dentro de un cierto límite es aceptable hasta 3 ó 4 veces el período de registro*.

Esta limitación en cuanto a la extrapolación, que en la generalidad de los casos es muy difícil de satisfacer, depende además de la utilización que se le vaya a dar al evento así encontrado, de la confianza que se tenga a la distribución de frecuencias determinados y a las características de la muestra que se esté analizando.

Algunas veces, con el objeto de ampliar la muestra de datos por analizar y hacer una mejor inferencia, se puede utilizar el criterio del método de la estación-año** o bien correlacionarla con otras muestras mayores existentes dentro de la zona hidrológica que se esté estudiando.

8.7 Análisis de gastos máximos anuales

Dentro del estudio de datos hidrológicos, uno de los más usuales es el análisis de gastos máximos anuales, pues está relacionado directamente con el diseño de vertedores de presas, capacidad de drenajes en caminos, obras de protección, encauzamiento y de defensa en ríos, etc.

* Ven Te Chow "Frecuency analysis of hydrologic date with special application to Rainfall Intensities". Bull. Universidad de Illinois, Vol 50, No. 81, Julio, 1958.

** Katharine Clarke-Hafstad, "Reliability of Station-Year Rainfall-Frequency Determinations", Trans ASCE, pags. 633, 1942.

El análisis de gastos es costumbre realizarlo utilizando las máximas anuales, aunque en ocasiones se utilizan los excedentes anuales (véase subinciso 8.4.1).

El análisis de gastos máximos anuales consiste en la obtención de su distribución de frecuencias, para posteriormente obtener a partir de ella el gasto de diseño para un cierto período de retorno.

Primeramente se obtienen los períodos de retorno ligado al registro disponible de gastos máximos anuales. Para esto, los gastos máximos anuales se ordenan en forma decreciente asignándoles su número de orden y con la ec (8.32) se determinan sus períodos de retorno correspondientes. De esta manera, se tiene para cada gasto máximo anual registrados o excedente anual, su períodos de retorno y por ende, su probabilidad de recurrencia (Ejemplo 8.11).

Conocida esta muestra de parejas de valores, se procede a determinar su distribución de probabilidades, seleccionando diversas distribuciones teóricas y probando cual es la más representativa de la muestra en estudio (inciso 8.1). Afortunadamente, se ha visto que en general los gastos máximos anuales, siguen alguna de las distribuciones indicadas en el inciso 8.5, lo que hace el cálculo más expedito. (ejemplos 8.12 y 8.13)

Determinada la distribución de frecuencias de los gastos máximos anuales que se esten analizando, se procede a valuar el período de retorno -

asignado al diseño (subinciso 8.4.4).

Obtenido el período de retorno de diseño, a partir de la distribución de gastos conocida, se obtiene el gasto máximo de diseño. Se debe tener en cuenta las limintación que implica una extrapolación cuando el período de retorno de diseño es grande comparado con los años de registro disponibles (inciso 8.6).

8.8 Análisis de lluvias

El estudio de frecuencias de lluvias es muy utilizado principalmente en hidrología en el estudio de predicción de tormentas, transporte de las mismas y su relación con escurrimientos, para el diseño de drenajes en caminos como en ciudades y aeropuertos.

El análisis de frecuencias de lluvia se aplica en forma directa a registros de pluviómetros si se estudian alturas de lluvias diarias y principalmente a pluviógrafos en donde se tiene un registro completo de todas las variables de las características fundamentales de la lluvia. Este análisis se puede hacer extensivo a áreas, utilizando algunas de las técnicas vistas en el inciso 3.4.

El proceso que se sigue en el análisis de lluvias puntuales es similar al descrito para el análisis de gastos, solo que en general se trabaja con series de datos excedentes anuales y con dos variables en lugar de una.

Estas dos variables comprenden las características que definen a una -precipitación pluvial, que son su altura de lluvia hp y su duración d. Por lo anterior, para definir una cierta distribución de probabilidades de alguna de las dos variables, se deberá considerar constantes una de ellas para definir a la otra. A partir de un análisis de este tipo, es factible correlacionar posteriormente las dos variables, involucrando sus frecuencias de incidencia.

8.8.1 Distribuciones de alturas de lluvia por una duración constante

Cuando se requeriere conocer la distribución de probabilidades de altura de lluvia máximas con cierta duración constante, se procede en forma similar al análisis indicado en el inciso 8.7, solo que sí se trabaja con excedentes anuales, para valuar el período de retorno se utiliza la ec. 8.33.

8.8.2 Curvas de alturas de lluvia-duración-período de retorno

La obtención de las curvas de altura ó intensidad de lluvia-duración-período de retorno de un registro de lluvia permite tener un conocimiento de la variación de las características de las lluvias en relación con sus frecuencias de incidencia.

Antes de proceder a utilizar una técnica de análisis, se requiere procesar y ordenar los datos disponivles. Si se observa un registro de pluviógrafo (fig 3.8), como este es continuo, existen múltiples combinaciones para asociar la altura de lluvia con una cierta duración que son valores discretos.

Una manera de proceder es considerar solo las características finales de las tormentas, relacionando su altura de lluvia total con su duración total. Esto implica, perder información, pues no se valúa la variación existente entre estas variables durante el proceso de la lluvia. Conforme aumenta el número de años de registro, la pérdida de información disminuye. Con este criterio, para utilizar la información obtenida se requiere agrupar las alturas en base a duraciones constantes. Para ello, se seleccionan intervalos de duración de lluvia constante, considerando representativos de cada uno de ellos, su duración media. De esta manera, se tendrá para cada duración característica de lluvia un grupo de alturas de lluvia.

Otra forma de análisis de los registros de lluvia correspondiente, es valuando la máxima variación de la altura de lluvia respecto a un intervalo de duración constante. Para esto, se requiere conocer la curva masa de cada tormenta que se esté estudiando. Conocida la curva masa, en un papel transparente se hacen divisiones verticales de las duraciones que estén analizando, generalmente múltiplos de un cierto intervalo de -

tiempo. Así, se procede a superponer el papel transparente en la curva masa de la tormenta que se esté analizando, y desplazando el origen de las abscisas, se determina el máximo incremento de altura de lluvia teniendo en un cierto intervalo de tiempo , fig 8.11.

De esta manera, se obtienen las condiciones más críticas de alturas de lluvia-duración- para cada tormenta. Observese que en este caso, terminando el proceso de análisis, se disponen de grupos de altura de lluvia para duraciones constantes.

Conocidas las características de las lluvias más desfavorables, se pueden utilizar varios criterios para obtener las curvas de altura o intensidad de lluvia-duración-período de retorno. A continuación se indican dos criterios a seguir.

8.8.2.1 Criterio propuesto por Chow.

Teniendo disponible la información de las características de las tormentas, agrupadas las alturas de lluvia para diferentes duraciones, Chow* considera que para cada duración constante, las alturas de lluvia correspondiente se les puede dar un tratamiento de series excedentes y aplicar algunos de los criterios vistos en el inciso 8.5. En este caso, el proceso es similar al empleado en el subinciso 8.7.1

De esta manera se tendrá, para cada duración constantes, una curva del

---

*Ven Te Chow "Frequency analysis of hydrologic date with special application to Rainfall Intensities" Bull. Universidad de Illinois, Vol 50, No. 81, julio 1953.

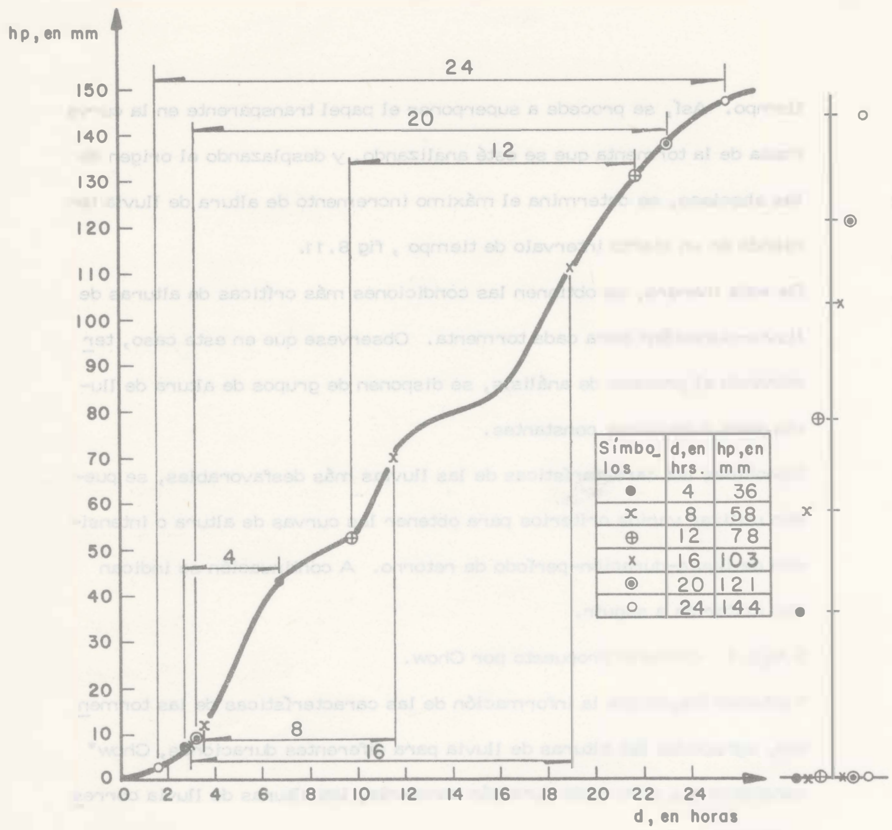


Fig.8.11 Análisis de una curva masa para obtener las máximas variaciones de altura de lluvia en relación con diversos intervalos de duración de la misma

tipo de la ec (8.44) de forma similar a

$$hp = a + b \log Te \qquad (8.46)$$

donde a y b son parámetros de ajuste, hp la altura de lluvia, en mm y Te el período de retorno, en años. Como esta expresión es para una duración constante, dividiendola entre ella, se obtiene una relación entre intensidades de lluvia y períodos de retorno, que es otra forma de expresar la ecuación anterior.

8.8.2.2 Ajuste por correlación lineal múltiple

Otra forma de valuar las curvas de altura de lluvia-duración-períodos de retorno, es obtener la ecuación de mejor ajuste entre los diversos grupos de valores de d, hp y t.

Para ello, antes de proceder a plantear el tipo de ecuación más conveniente, es necesario tener una idea de cuales son los tipos de ecuaciones que en general siguen estos valores.

Dentro de las ecuaciones más usuales, se tiene la del tipo de:

$$i = \frac{K T^h}{d^n} \qquad (8.47)$$

donde K, h y n son parámetros, d la duración de la lluvia, i su intensidad y T su período de retorno.

Tomando logaritmos a esta ecuación, se obtiene que

$$\log i = \log K + h \log T - n \log d \qquad (8.48)$$

y transformando, esta ecuación puede escribirse como

$$y = a_0 + a_1\, x_1 + a_2\, x_2$$

donde $\log i = y$, $\log T = x_1$ y $\log k = x_2$

Esta ecuación es similar a la (8.16), cuya obtención de los parámetros a_0, a_1 y a_2 se indican en el inciso 8.3.

De esta manera, se puede obtener para la agrupación de las características de las lluvias que se estén estudiando una ecuación del tipo de la ec (8.47).

Ejemplo 8.14   En la tabla 8.23 se proporcionan los datos de las intensidades máximas de lluvia registradas en la Estación Sta. Catarina, en el estado de Nuevo León. Se desea obtener en base a estos registros las curvas de altura de precipitación-duración-período de retorno - (hp-d-T) para dicha estación aplicando:

a) Criterio de Chow

b) Ajuste por Correlación Lineal Múltiple

De la Tabla 8.23 primeramente como los datos se encuentran en intensidad de lluvia $i = hp / d$, habrá que obtener para cada valor de la intensidad su correspondiente altura de precipitación "hp" en relación con su - duración; hecho esto para cada duración "d", se ordenan los valores en forma decreciente de mayor a menor, y se obtiene para cada duración -

TABLA 8.23

INTENSIDADES MAXIMAS DE LLUVIA EN LAS CUENCAS DEL RIO SAN JUAN DEL BAJO BRAVO Y DE LA REGION GOLFO NORTE

PRECIPITACIONES EN MILIMETROS POR HORA

FECHAS			MINUTOS									
			5	10	15	20	30	45	60	80	100	120
SANTA CATARINA NUEVO LEON												
1938	Feb	20	120.0									
	Jul	29	120.0	114.0	97.2	87.2	80.0	63.3	52.5	41.4	33.5	28.0
1939	Abr	12	132.0	109.8	84.0	80.1	58.0	40.5	31.5	24.1	19.3	16.1
1940	Jun	24		64.2	49.2	43.2						
	Ago	9	96.0				40.0	37.6	29.6	21.9	17.5	14.6
1941	May	5	79.2	52.2	39.6	31.5	21.0	14.0	10.5	7.9		
	Jun	9									7.7	6.9
1942	Jul	4		96.0	80.0	77.7	72.0	74.0	66.0	50.1	40.3	33.9
	Jul	5	148.8		80.0							
1943	Sep	6	126.0	76.2	59.2	48.3	34.4	23.1	17.4	14.6	14.4	12.6
1944	Oct	7	92.4	63.6	56.0	48.6	40.0	34.7	30.0	24.2	23.4	23.0
1945	Ago	30	86.4	61.8	51.2	42.6	36.0	26.7	24.7	24.0	23.4	22.3
1946	Ago	30	102.0	58.2	50.4	45.0	31.4	21.1	15.8	11.9	9.5	7.5
1947	Jul	30		60.0	56.0	51.3	42.6	31.3	24.0			
	Ago	4								21.5	20.4	17.9
	Ago	25	120.0	60.0								
1948	Jul	7	76.8	57.6	44.0	35.1	25.0	24.7	20.0			
	Sep	9							20.0	16.7	15.8	13.1
1949	Sep	19	98.4	57.0	46.8	54.0	40.0	30.7	30.0	25.5	22.5	19.3
1950	Mar	3							6.8	6.5	5.6	4.7
	Jul	13			23.0	18.3	12.6	8.4				
	Ago	18	57.6	28.8								
1951	Jun	24	128.4	93.0	80.0	85.0	62.2	47.3	36.4	27.3	21.8	18.2
1952	Abr	23	66.0									
	Jun	7		46.8	34.0	27.0	18.4	12.7	10.0	7.5	6.7	5.9

TABLA 8.23 (Continuación)

INTENSIDADES MAXIMAS DE LLUVIA EN LAS CUENCAS DEL RIO SAN JUAN DEL BAJO BRAVO Y DE LA REGION GOLFO NORTE

PRECIPITACIONES EN MILIMETROS POR HORA

FECHAS			MINUTOS									
			5	10	15	20	30	45	60	80	100	120
1953	Jul	14	120.0				40.0	40.0				
	Oct	3	120.0	67.8	56.0	48.6	40.0	40.0	35.0	28.5	22.8	19.0
1954	Oct	5						14.0	12.0	9.6	8.6	7.1
	Oct	8	96.0	54.0	37.2	27.9	18.6					
1955	Jul	8	96.0	48.0								
	Nov	2		48.0	48.0	43.5	37.0	27.3	27.5	25.5	24.0	24.0
1956	May	15	150.0	93.0	76.0	60.0	41.0	33.0	25.5	19.1	15.3	12.8
1957	Sep	21	90.0	66.0	48.0	42.9	38.0	25.3	21.6	19.3	16.0	14.5
						SIN DATOS						
1959	Jun	14	68.4		36.0	27.6	18.6	13.3	11.4	11.4	9.1	7.8
	Ago	13		40.8								
1960	Ago	11	117.6	70.2	60.0	54.0	40.0	27.4	20.6	15.8	12.9	11.3
1961	Jul	10	85.2	42.6	28.4	21.3	14.2	9.4	7.1	5.3	4.3	3.6
1962	Sep	10	162.0	111.0	80.0	62.1	60.0	51.3	43.5	45.0	36.0	40.0
1963	May	17	96.0	60.0	40.0	34.5					16.3	15.0
	Jun	16					27.0	27.0	22.3	17.3	16.3	
1964	May	31	120.0	105.0	71.2	53.4	35.6	24.7	17.8	13.4	10.7	8.9

su período de retorno asociado "Te" , que en términos de series de exce-

dentes anuales resulta Te = n/m (ec. 8.33) n = 26

ejemplos: "hp" de la tabla - dato

1.- d = 5 min 1938: $i = 120 \ \frac{mm}{hr}$ tabla - cálculo

d = 5 min hp = 120(5)/60 = 10.0 mm; con orden 7o.

2.- d = 80 min 1942: i = 50.1 mm/hr d = 80 min

hp = 50.1 (80)/60 = 66.8 mm; con orden 1o. "Te"

3.- orden = 1 = m; n = 26

$$T_e = \frac{n}{m} = \frac{26}{1} = 26$$

4.- orden = 17 = m

$$Te = \frac{n}{m} = \frac{26}{17} = 1.5294 = 1.53$$

En la Tabla 8.24 se muestra la obtención y ordenación de los datos para su aplicación.

a) Criterio de Chow

1. Relación para una duración de lluvia d = ctte = 5 min

$$hp = a + b \ \log \ Te$$

de acuerdo ec (8.13)

$$y' = a + b \ x_1$$

donde $y' = hp$

$$x_1 = \log \ Te$$

TABLA 8.24 OBTENCION Y ORDENACION DE VALORES

Orden (m)	d min / Te años	5	10	15	20	30	45	60	80	100	120
1	26.00	13.5	19.0	24.3	29.0	40.0	55.5	66.0	66.8	67.2	80.0
2	13.00	12.5	18.5	21.0	28.5	36.0	47.5	52.5	60.0	60.0	67.8
3	8.67	12.4	18.3	20.0	26.7	31.1	38.5	43.5	55.2	55.8	56.0
4	6.50	11.0	17.5	20.0	25.9	30.0	35.5	36.4	38.0	40.0	48.0
5	5.20	10.7	16.0	20.0	20.7	29.0	30.4	35.0	36.4	39.0	46.0
6	4.33	10.5	15.5	19.0	20.0	21.3	30.0	31.5	34.0	39.0	44.6
7	3.71	10.0	15.5	17.8	18.0	20.5	28.2	30.0	34.0	38.0	38.6
8	3.25	10.0	12.7	15.0	18.0	20.0	26.0	30.0	32.3	37.5	38.0
9	2.89	10.0	11.7	14.8	17.8	20.0	24.8	29.2	32.1	36.3	36.4
10	2.60	10.0	11.3	14.0	17.1	20.0	23.5	27.5	32.0	34.0	35.8
11	2.36	9.8	11.0	14.0	16.2	20.0	23.0	25.5	29.2	32.2	32.2
12	2.17	8.5	10.7	14.0	16.2	20.0	20.6	24.7	28.7	29.2	30.0
13	2.00	8.2	10.6	12.8	16.1	19.0	20.5	24.0	25.7	27.2	29.2
14	1.86	8.0	10.3	12.6	15.0	18.5	20.3	22.3	25.5	26.7	29.0
15	1.73	8.0	10.0	12.3	14.5	18.0	20.0	21.6	23.1	26.3	26.2
16	1.63	8.0	10.0	12.0	14.4	17.8	19.0	20.6	22.3	25.5	25.6
17	1.53	8.0	9.7	12.0	14.3	17.2	18.5	20.0	21.1	24.0	25.2
18	1.44	7.7	9.6	11.7	14.2	15.7	18.5	17.8	19.5	21.5	22.6
19	1.37	7.5	9.5	11.0	11.7	13.5	17.3	17.4	17.9	17.8	17.8
20	1.30	7.2	9.0	10.0	11.5	12.5	15.8	15.8	15.9	15.8	15.8
21	1.24	7.1	8.7	9.9	10.5	10.5	10.5	12.0	15.2	15.2	15.6
22	1.18	6.6	8.0	9.3	9.3	9.3	10.5	11.4	10.8	13.8	14.2
23	1.13	6.4	7.8	9.0	9.2	9.3	10.0	10.5	10.5	12.8	13.8
24	1.08	5.7	7.1	8.5	9.0	9.2	9.5	10.0	10.0	11.2	11.8
25	1.04	5.5	6.8	7.1	7.1	7.1	7.1	7.1	8.7	9.3	9.4
26	1.00	4.8	4.8	5.9	6.1	6.3	6.3	6.8	7.1	7.2	7.2

y se trata de una correlación lineal simple; aplicando el criterio expuesto en la sección de la Tabla 8.25 se tiene que:

$$n \sum_{i=1}^{n} x_1^2 = 26\ (7.2966) = 189.7116$$

$$(\sum_{i=1}^{n} x_1)^2 = (10.1821)^2 = 103.6752$$

$$n \sum_{i=1}^{n} x_1 y' = 16\ (108.5597) = 2822.5522$$

$$n \sum_{i=1}^{n} (y')^2 = 26 (2117.62) = 55058.12$$

$$(\sum_{i=1}^{n} y')^2 = (227.6)^2 = 51801.76$$

Con lo cual substituyendo en las ecs (8.17) y (8.18) se deduce

$$Sxx = 189.7116 - 103.6752 = 86.0364$$

$$Sxy = 2800.5522 - 2317.4460 = 505.1062$$

$$Syy = 55058.12 \quad -51801.76 \quad = 3256.36$$

siendo de la ec (8.15)

$$b = \frac{505.1062}{86.0364} = 5.8708 = 5.87$$

$$\bar{y}' = \frac{227.60}{26} = 8.754$$

TABLA 8.25 hp = f (Te) para d = 5 min.

Orden (m)	Y'	Te	$X_1 = \log Te$	$(X_1)^2$	$(Y')^2$	$X_1 Y'$
1	13.5	26.00	1.4150	2.0022	182.25	19.1025
2	12.5	13.00	1.1139	1.2407	156.25	13.9237
3	12.4	8.67	0.9380	0.8789	153.76	11.6312
4	11.0	6.60	0.8129	0.6608	121.00	8.9419
5	10.7	5.20	0.7160	0.5126	114.49	7.6612
6	10.5	4.33	0.6365	0.4051	110.25	6.6830
7	10.0	3.71	0.5694	0.3242	100.00	5.6940
8	10.0	3.25	0.5119	0.2620	100.00	5.1190
9	10.0	2.89	0.4609	0.2124	100.00	4.6090
10	10.0	2.60	0.4150	0.1722	100.00	4.1500
11	9.8	2.36	0.3729	0.1390	96.04	3.6544
12	8.5	2.17	0.3365	0.1132	72.25	2.8602
13	8.2	2.00	0.3010	0.0906	67.24	2.4682
14	8.0	1.86	0.2695	0.0726	64.00	2.1560
15	8.0	1.73	0.2380	0.0566	64.00	1.9040
16	8.0	1.63	0.2122	0.0450	64.00	1.6976
17	8.0	1.53	0.1847	0.0341	64.00	1.4776
18	7.7	1.44	0.1584	0.0250	59.29	1.2196
19	7.5	1.37	0.1367	0.0186	56.25	1.0252
20	7.2	1.30	0.1139	0.0129	51.84	0.8200
21	7.1	1.24	0.0934	0.0087	50.41	0.6631
22	6.6	1.18	0.0719	0.0051	43.56	0.4745
23	6.4	1.13	0.0531	0.0028	40.96	0.3398
24	5.7	1.08	0.0334	0.0011	32.49	0.1903
25	5.5	1.04	0.0170	0.0002	30.25	0.0935
26	4.8	1.00	0.000	0.0000	23.04	0.0000
SUMA	227.60		10.1821	7.2966	2117.621	108.5597

$$\bar{x} = \frac{10.1821}{26} = 0.392$$

y de la ec (8.16)

$$a = 8.754 - 5.8708\,(0.392) = 6.455$$

la ecuación buscada para d = 5 minutos = ctte es:

$$y' = a + b\,x_1$$

$$x_1 = \log\ Te \quad ; \quad y' = hp$$

$$hp = 6.455 + 5.87 \log Te \qquad (1)$$

para d = 5 min

y su coeficiente de correlación

$$rxy = \frac{505.1062}{\left[(86.0364)\,(3256.36)\right]^{1/2}} = 0.9543$$

analogamente para las distintas duraciones se tiene:

d = 10 min

$$hp_{d=10} = 7.473 + 10.343 \log Te \qquad (2)$$
$$r = 0.9578$$

d = 15 min $\quad hp_{d=15} = 8.885 + 12.470 \log Te \qquad (3)$
$$r = 0.9636$$

d = 20 min $\quad hp_{d=20} = 9.517 + 16.650 \log Te \qquad (4)$
$$r = 0.9647$$

d = 30 min $\quad hp_{d=30} = 9.942 + 22.914 \log Te \qquad (5)$
$$r = 0.9672$$

d = 45 min $\quad hp_{d=45} = 9.935 + 32.312 \log Te \qquad (6)$
$$r = 0.9859$$

$\bar{x}$

d = 60 min	$hp_{d=60} = 10.217 + 37.660 \log Te$ $r = 0.9865$	(7)
d = 80 min	$hp_{d=80} = 11.365 + 41.101 \log Te$ $r = 0.9796$	(8)
d = 100 min	$hp_{d=100} = 13.103 + 41.429 \log Te$ $r = 0.9755$	(9)
d = 120 min	$hp_{d=120} = 12.327 + 48.744 \log Te$ $r = 0.9903$	(10)

analizando apriori los resultado, se observa que el ajuste realizado es bueno dada la magnitud del índice de correlación o coeficiente. En función de las relaciones obtenidas hp = f (Te) para una duración de lluvia constante, podrían obtenerse en la tabla 8.23 los valores correspondiente al año de 1958, interpolando el valor del período de retorno para los años adyacentes 1957 y 1959 para su correspondiente duración; así tendríamos:

para d = 5 min

i_{57} = 90 mm/hr hp_{57} = 7.5 mm Te_{57} = 1.37 años

i_{59} = 68.4 mm/hr hp_{59} = 5.7 mm Te_{59} = 1.08 años

El Te probable para el año de 1958

será $Te_{58} = (1.37 + 1.08) /2 = 1.23$ años

Substituyendo en la ecuación 1:

$$hp_{58} = 6.455 + 5.87 \log Te = 6.455 + 5.87 \log 1.23 =$$

$$= 6.455 + 5.87\ (0.0899) = 6.455 + 0.5277 = 6.9827$$

$$hp_{58} = 7.0 \text{ mm}$$

para $d = 5$ min

de forma análoga podemos obtener los restantes valores de hp para cada duración:

duración	5	10	15	20	30	45	60	80	100	120	minutos
Te	1.23	1.70	1.33	1.33	1.57	1.38	1.46	1.62	1.55	1.55	años
ecuación	1	2	3	4	5	6	7	8	9	10	
hp	7.0	9.9	10.4	11.6	14.4	14.5	16.4	20.0	21.0	21.6	mm

Si las ecuaciones obtenidas desean expresarse como una relación ya nó de la altura de precipitación hp, sino de la intensidad de lluvia "i" como función del período de retorno asociado "Te", bastará dividir la altura de precipitación entre la duración para la cual fue deducida; por ejemplo la ec 1.-) $hp_{d=5} = 6.455 + 5.87 \log Te$ (mm)

dividiendo hp/5min ≙ hp/0.083 = 1 (mm/hora)

$$i = 77.46 + 70.44 \log Te \text{ en mm/5min.}$$

b) Ajuste por Correlación Lineal Múltiple

Se trata entonces de obtener una ecuación del tipo(8.47)

$$i = \frac{KT^h}{d^m}$$ es intensidad o bien en relación con la altura de precipitación "hp":

$$hp = \frac{KTe^h}{d^m}$$ con el método de correlación lineal múltiple inciso (8.3) que ajusta una relación del tipo

$$y = a_0 + a_1 x_1 + a_2 x_2 \qquad (3 \text{ variables})$$

Tomando logaritmos se tiene una ecuación de la forma (8.48).

$$\log hp = \log K + h \log Te + n \log d$$

siendo $y = \log hp \qquad a_0 = \log K$

$x_1 = \log Te \qquad a_1 = g$

$x_2 = \log d \qquad a_2 = n$

Para la estimación de los parámetros a_0, a_1, a_2 se emplea el análisis de correlación lineal de 3 variables, el que arroja de su planteamiento matricial, el sistema de ecuaciones:

$$y = a_0 n + a_1 \sum x_1 + a_2 \sum x_2$$
$$yx_1 = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 \quad x_2$$
$$yx_2 = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 \; . \; . \qquad (I)$$

Para su aplicación se deberá procesar la información disponible, en este caso la mostrada en la Tabla 8.24. A fin de simplificar y por mayor claridad, el sistema anterior se resolvió solo con 10 grupos de valores, la cual se muestra en la Tabla 8.25.

TABLA 8.26 ORDENACION Y CALCULO

Orden	hp (mm)	Te (años)	d (min)	y log hp	X_1 log Te	X_2 log d	$(X_1)^2$	$(X_2)^2$	$(X_1)(X_2)$	YX_1	YX_2
1	13.5	26.00	5	1.1303	1.4150	0.6990	2.0022	0.4886	0.9890	1.599	0.790
2	18.5	13.00	10	1.2672	1.1139	1.0000	1.2407	1.0000	1.1139	1.411	1.267
3	20.0	8.67	15	1.3010	0.9380	1.1761	0.8789	1.3832	1.1031	1.220	1.530
4	25.9	6.50	20	1.4133	0.8129	1.3010	0.6608	1.6926	1.0575	1.148	1.838
5	29.0	5.20	30	1.4624	0.7160	1.4771	0.5126	2.1818	1.0576	1.047	2.160
6	30.0	4.33	45	1.4771	0.6365	1.6532	0.4051	2.7330	1.0522	0.940	2.441
7	30.0	3.71	60	1.4771	0.5694	1.7782	0.3242	3.1619	1.0125	0.841	2.626
8	32.3	3.25	80	1.5092	0.5119	1.9031	0.2620	3.6217	0.9741	0.772	2.872
9	36.3	2.89	100	1.5599	0.4609	2.0000	0.2124	4.0000	0.9218	0.718	3.119
10	35.8	2.60	120	1.5539	0.4150	2.0792	0.1722	4.3230	0.8628	0.644	3.230
SUMA				14.1514	7.5895	15.0669	6.6711	24.5858	10.1445	10.340	21.873
				y	X_1	X_2	X_1^2	X_2^2	$X_1 X_2$	yX_1	yX_2

sustituyendo los valores en el sistema (I)

$$14.1514 = 10\,a_0 + 7.5895\,a_1 + 15.0669\,a_2$$

$$10.3400 = 7.5895\,a_0 + 6.6711\,a_1 + 10.1445\,a_2$$

$$21.8730 = 15.0669\,a_0 + 10.1445\,a_1 + 24.5858\,a_2$$

resolviendo simultáneamente este sistema de ecuaciones obtenidas

$a_0 = 2.4645 = \log k$ $\qquad K = \text{antilog } a_0$

$a_1 = -0.8301$ $\qquad K = 291.4$

$a_2 = -0.2769$

la ec queda:

$$y_0 = 291.4 - 0.8301 x_1 - 0.276\, x_2$$

$$\log hp = \log K + h \log Te + n \log d$$

$$\log K = 2.4645$$

$$h = a_1 = -0.8301$$

$$n = a_2 = -0.2769$$

subs: $\log hp = 2.4645 - 0.8301 \log Te - 0.2769 \log d$

tomando antilogarítmos:

$$hp = \frac{291.40}{(Te)^{0.83}\,(d)^{0.28}} \quad \ldots (A)$$

Tomando otra serie de datos de la Tabla 8.12 a se puede construir la Tabla 8.27.

$$14.30 = 10\ a_0 + 7.59\ a_1 + 15.07\ a_2$$

$$11.71 = 7.59\ a_0 + 6.67\ a_1 + 12.62\ a_2$$

$$22.80 = 15.07\ a_0 + 12.62\ a_1 + 24.59\ a_2$$

se obtiene $a_0 = 0.513 = \log k\ ;\ \ k = \text{antilog}\ a_0 = 3.258$

$$a_1 = 427 = h$$

$$a_2 = 0.398 = n$$

$$y_0 = a_0 + a_1\ x_1 + a_2\ x_2$$

$$\log hp = \log k + h \log Te - n \log d$$

$$\log hp = 0.513 + 0.427 \log Te + 0.398 \log d$$

tomando antilogs:

$$hp = 3.258\ (Te)^{0.43}\ (d)^{0.40} \quad \ldots (B)$$

que es la relación pedida obteniendo algunos valores de "hp" con las relaciones encontradas A y B para comparación con los valores reales

hp real (mm)	Te años	d min	hp calc. A	hp calc. B
44.6	4.33	120	22.94	40.95
60.0	13.00	80	10.30	55.72
24.7	2.17	60	49.30	23.14
30.0	4.33	45	30.09	27.71
10.0	1.30	20	100.00	11.99
19.0	26.00	10	10.30	32.75

TABLA 8.27 ORDENAMIENTO DEL CALCULO

Orden (h)	hp (mm)	Te (años)	d (min)	y log hp	x_1 log Te	x_2 log d	$(x_1)^2$	$(x_2)^2$	$(x_1)(x_2)$	$Y x_1$	$Y x_2$
1	80.0	26.00	120	1.9031	1.4150	2.0792	2.0022	4.3230	2.9420	2.693	3.957
2	60.0	13.00	100	1.7780	1.1139	2.0000	1.2407	4.0000	2.2278	1.981	3.556
3	55.2	8.67	80	1.7419	0.9380	1.9031	0.8789	3.6217	1.7851	1.634	3.315
4	36.4	6.50	60	1.5611	0.8129	1.7782	0.6608	3.1619	1.4454	1.269	2.776
5	30.4	5.20	45	1.4829	0.7160	1.6532	0.5126	2.7330	1.1836	1.062	2.452
6	21.3	4.33	30	1.3284	0.6365	1.4771	0.4051	2.1818	0.9401	0.846	1.962
7	18.0	3.71	20	1.2553	0.5694	1.3010	0.3242	1.6926	0.7407	0.715	1.633
8	15.0	3.25	15	1.1761	0.5119	1.1761	0.2620	1.3832	0.6020	0.602	1.383
9	11.7	2.89	10	1.0682	0.4609	1.0000	0.2124	1.0000	0.4609	0.492	1.068
10	10.0	2.60	5	1.0000	0.4150	0.6990	0.1722	0.4886	0.2900	0.415	0.699
SUMA				14.2952 Σy	7.5895 Σx_1	15.0669 Σx_2	6.6711 $\Sigma (x_1)^2$	24.5858 Σx_2^2	12.6176 $\Sigma (x_1)(x_2)$	11.709 $\Sigma (y)(x_1)$	22.801 $\Sigma (y)(x_2)$

De la comparación de los valores calculados con los reales se puede concluir que la relación B proporciona un mejor ajuste a los datos reales, por tanto la ecuación final es

$$hp = 3.258 \ (Te)^{2.43} \ (d)^{0.40}$$

la cual se muestra en forma gráfica en la fig.8.12. Con esta ecuación se deduce para el año de 1958 la siguiente información:

DATOS FALTANTES PARA EL AÑO 1958 (deducidos)

duración	5	10	15	20	30	45	60	80	100	120	minutos
Te	1.23	1.70	1.33	1.33	1.57	1.38	1.38	1.60	1.55	1.55	años
hp	6.8	20.0	10.8	18.1	15.3	17.0	19.5	22.9	24.6	26.5	mm

8.8.3 Ajustes y utilización de las curvas de altura de lluvia - duración - período de retorno.

Cuando no se dispone de la información suficiente para poder obtener en una estación determinada, sus curvas de altura de lluvia - duración - período de retorno, y se requiere analizar lluvias con duraciones hasta de 120 minutos, se pueden aplicar las ecuaciones propuestas por Bell*. Estas ecuaciones permiten obtener la altura de lluvia para una cierta duración entre 5 y 120 min. y período de retorno, entre dos y cien años, si se conoce la altura de lluvia con duración de 60 min. y período de retorno de 2 ó 10 años.

---

* Frederick Charles Bell "Generalized Rainfall-Duration-Frequency Relationships", Journal of the Hydraulics Division, Proceedings of the A.S.C.E. Enero 1969.

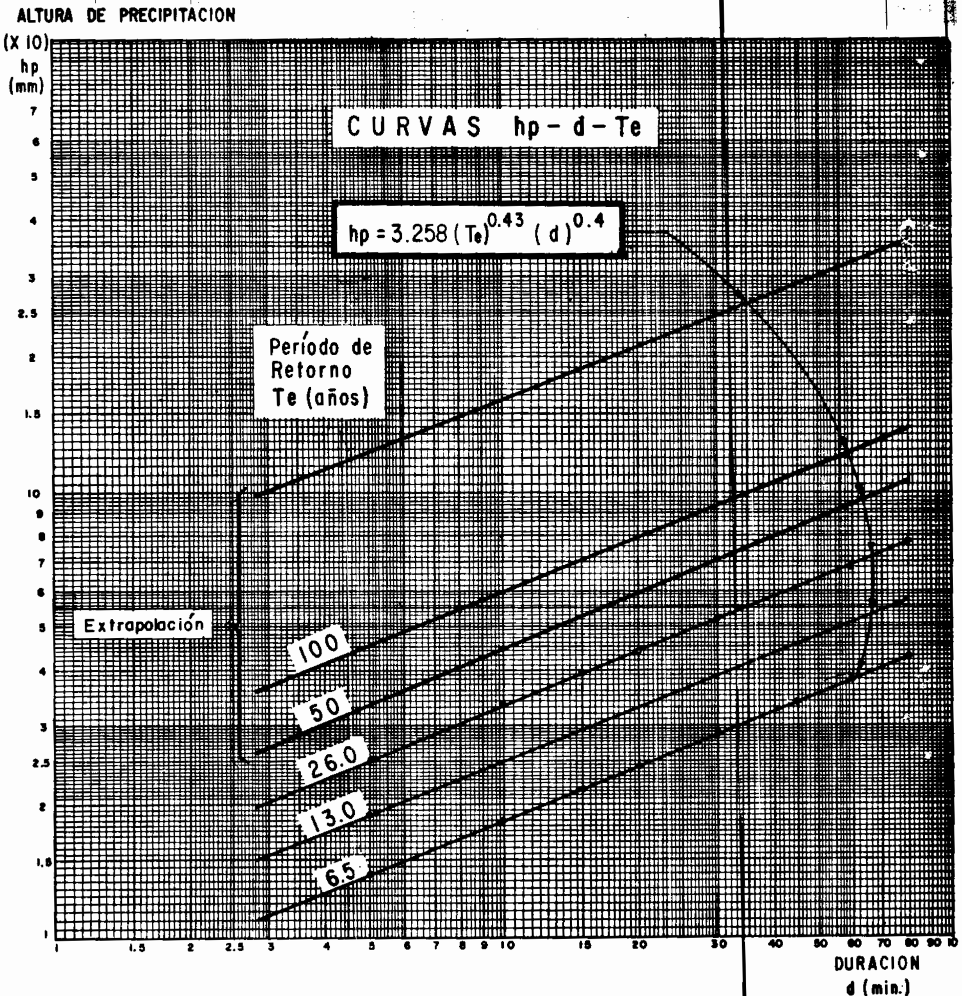


Fig. 8.12 Curvas "hp-d-Te" (mm)(min)(años) para la estación Sta. Catarina, N. L.

Las ecuaciones son:

$$hp = (0.35 \log_e T + 0.76)(0.54\, d^{0.25} - 0.50)\, hp' \quad (8.50)$$

$$hp = (0.21 \log_e T + 0.52)(0.54\, d^{0.25} - 0.50)\, hp'' \quad (8.51)$$

donde

d	**duración de la lluvia, entre 5 y 120 min.**
hp	altura de lluvia para una cierta duración y período de retorno, en mm.
hp'	altura de lluvia para duración de 60 min. y período de retorno de 2 años, en mm
hp''	altura de lluvia para una duración de 60 min. y período de retorno de 10 años, en mm

Para realizar aplicaciones de las curvas de altura de lluvia-duración-período de retorno, se requiere hacerlas extensivas a áreas. Para ello, se pueden utilizar algunos de los criterios vistos en el subinciso 3.4.1 y el subinciso 3.5.2, considerando períodos de retornos constantes. Otra forma es en base a una relación lineal, aceptando un plano de isoyetas de determinadas características, por ejemplo, de altura de lluvia media anual fig (3.16) o bien, de alturas de lluvia para una cierta duración y período de retorno, tal que sea representativa de las lluvias en la región. De esta manera, las alturas de lluvia de las curvas de altura de lluvia-duración-período de retorno se ajustan con una relación lineal, entre la altura de lluvia media deducida en el área en la cual se quiera hacer exten sivas las curvas y la altura correspondiente al punto donde se efectuó la obtención de dichas curvas.

Una aplicación directa de las curvas ajustadas de altura de lluvia-duración-período de retorno es la obtención de hietogramas de lluvias para una cierta cuenca, las cuales posteriormente, y en base a algún modelo de lluvia-escurrimiento, permiten la obtención de avenidas o gastos de diseño. Para poder utilizar las curvas antes mencionadas, se requiere valuar primero el período de retorno de diseño (subinciso 8.4.4).

Asignado un período de retorno, las curvas se transforman en una relación entre alturas de lluvia y duraciones, fig (8.12). Para fijar la duración de la tormenta es costumbre considerarla igual al tiempo de concentración del escurrimiento, aceptando apriori que la lluvia así deducida tiene distribución uniforme en la cuenca donde se está aplicando y proporciona las condiciones de escurrimiento más desfavorables. En el siguiente capítulo se plantean diversas ecuaciones para valuar los tiempos antes mencionados, así como la forma de proceder para obtener estas ecuaciones.

Aceptando que se conoce la duración de la tormenta, directamente de la relación antes mencionada se puede valuar su altura de lluvia. Para obtener el hietograma correspondiente, la duración de la tormenta se subdivide en múltiplos de un cierto intervalo constante, obteniendo para cada múltiplo, una altura de lluvia total. La altura de lluvia de cada intervalo será la diferencia entre la altura de lluvia obtenida al inicio y final del intervalo. Debido al proceso de análisis, el hietograma así deducido tendrá una variación lineal de las alturas de lluvia siendo máxima al principio y mínima al final. Tratando de ser más congruentes con la realidad, conviene distribuir los incrementos de lluvia, sin perder su liga, de tal manera que máximo incremento esté a la mitad o un tercio de la duración total de la lluvia.

Ejemplo 8.15 Se desea obtener el hietograma de la tormenta con duración de 6 hrs. y para un período de retorno Te = 100 años, a partir de la ecuación de la curva "hp-k-Te" deducida para la estación Santa Catarina en el Edo. de Nuevo León del ejemplo 8.12. De estudios efectuados en la región de interés se conoce que la relación entre la precipitación media respecto a la precipitación registradora en la estación es de 0.76.

Del ejemplo anterior se tiene que para la Estación Santa Catarina, el comportamiento de la lluvia se puede representar como:

$$hp = 3.258\,(Te)^{0.427}\,(d)^{0.40}$$

De acuerdo con la información disponible, la ecuación anterior se transforma a:

$$hpm = 3.258\,(0.76)\,(Te)^{0.427}\,(d)^{0.40}$$

finalmente la ecuación de la que obtendremos el hietograma para Te = 100 años y d = 6 hrs.

$$hpm = 2.476\ (Te)^{0.427}\ (d)^{0.40} \quad \text{ó bien}$$

$$\log hpm = \log 2.476 + 0.427 \log Te + 0.40 \log d$$

Te = 100 años $\log hpm = 0.394 + 0.427 \log Te + 0.40 \log d$

T A B L A 8.28

OBTENCION DEL CALCULO (Ejemplo 8.15)

d (hr)	d (min)	log d	log hpm (acum)	d (hr)	d (hrs)	hpm acum	hpm (mm)
6	360	2.5563	2.2705	1	5 - 6	186.1	13.5
5	300	2.4771	2.2388	1	4 - 5	172.6	14.5
4	240	2.3802	2.2000	1	3 - 4	158.1	16.9
3	180	2.2553	2.1501	1	2 - 3	141.2	21.4
2	120	2.0792	2.0796	1	1 - 2	119.8	28.9
1	60	1.7780	1.9592	1	0 - 1	90.9	90.9

SUMA = 186.1

Como el análisis se hace considerando duraciones totales es conveniente distribuir las variaciones de las alturas de lluvia obtenidas, teniendo en cuenta que la máxima variación generalmente se presenta a 1/3 ó 1/2 de la duración total. De acuerdo con esto, en la fig 8.13 se muestra el hietograma resultante.

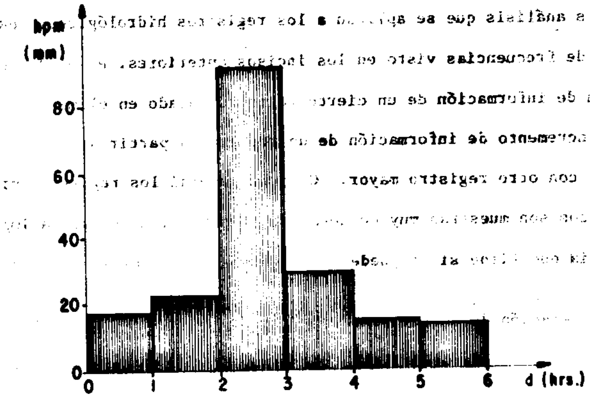


Fig. 8.13 Hietograma de la precipitación media deducida de las curvas "hp-d-Te" para Te = 100 años.

8.9 Generación de Información Hidrológica

Otros análisis que se aplican a los registros hidrológicos, aparte del de frecuencias visto en los incisos anteriores, es el de generación de información de un cierto registro basado en el mismo y el de incremento de información de un registro a partir de su correlación con otro registro mayor. Como en general los registros hidrológicos son muestras muy pequeñas de su población, se ve la importancia que tiene si se puede incrementar la información disponible.

La generación de un registro hidrológico, se aplica fundamentalmente a volúmenes de escurrimientos, con lo cual es posible realizar funcionamientos de vasos de almacenamiento en presas, así como el dimensionamiento de éstos. Por otra parte, el poder incrementar la información de un registro a partir de otro, complementa la información disponible para un mejor análisis de frecuencias ó de generación del mismo. Estos criterios también se aplican para generar volúmenes de escurrimiento a partir del comportamiento de las lluvias. Las técnicas de generación se basan en el hecho de que un registro de datos hidrológicos, es una secuencia ordenada de valores respecto al tiempo y es factible darles un tratamiento de series cronológicas

8.9.1 Series Cronológicas

La mayoría de los métodos estadísticos usados en el análisis de series de datos hidrológicos, consideran que las series son estacionarias y que las observaciones, son distribuidas independientemente en el tiempo. Esto implica por una parte, que los parámetros estadísticos de la muestra se conserven y por otra, que la ocurrencia de un evento se considere independiente de todos los eventos previos. Aunque el hecho de que la serie sea estacionaria en general se acepta como tal por la naturaleza de los registros, la dependencia entre las observaciones hidrológicas decrece con el incremento en su tiempo-base; así, los gastos diarios no son distribuídos independientemente en el tiempo, la dependencia entre los gastos mensuales y los gastos anuales es menor que entre los gastos mensuales.

El primer paso para analizar una serie cronológica es separar los diversos elementos de que esta compuesta, para posteriormente, si la serie sigue un cierto proceso, determinarlo y obtener el proceso generativo de la misma.

Una serie cronológica de datos hidrológicos puede considerarse compuesta por la suma de un elemento no aleatorio y un elemento aleatorio. Un elemento no aleatorio es aquél que se tiene cuando los valores separados K unidades de tiempo.

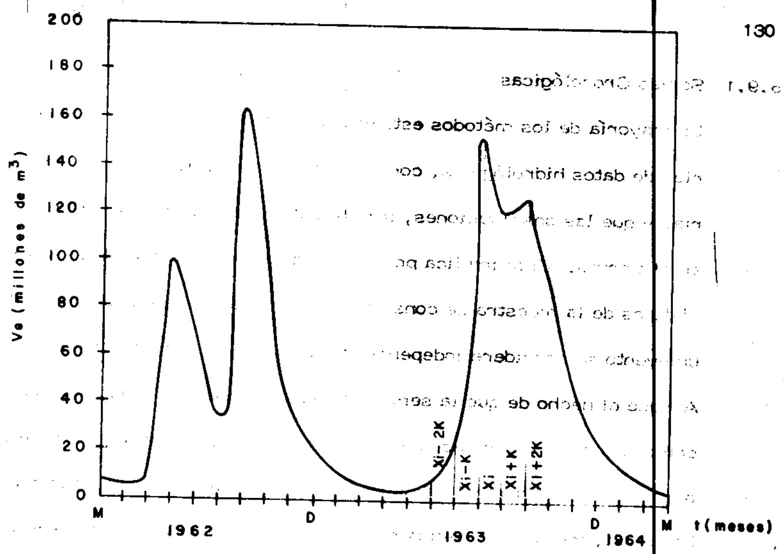


Fig. 8.14 Registro de volumenes de escurrimientos mensuales del Río Frío en la estación del mismo nombre

Si los valores de una serie cronológica se les denomina con x_i, (fig.8.14) si son linealmente dependientes de las observaciones desplazadas k unidades de tiempo, o sea de las $x_i + k$, la correlación entre x_i y $x_i + k$ se puede tomar como una medida de su dependencia y se le denomina correlación seriada de orden k. Este coeficiente de correlación seriado es análogo al coeficiente de correlación para dos grupos de datos visto en el inciso 8.2, ec 8.19, solo que en este caso x_i y $x_i + k$ se consideran los dos grupos de datos. En este caso para $k = 0$, $r_0 = 1$ y si la serie no es aleatoria para $k \geq 1$, $-1 < r_k < 1$. Si una serie cronológica es aleatoria $r_k = 0$ para todos los valores de $k \geq 1$.

Los elementos no aleatorios de una serie pueden estar compuestos tanto por una tendencia ó un movimiento en largos períodos y una oscilación sobre la tendencia, o solo por alguno de los dos. La tendencia es un movimiento lento de una serie en un largo período de tiempo; para definirla con precisión en una serie, se requiere tener el registro completo de ésta. En general, como lo anterior no es posible, en las series hidrológicas no se toma en cuenta, o bien se considera como la media de todos los valores por analizar, lo cual implícitamente es lo mismo.

En cuanto a movimientos oscilatorios, en una serie de datos hidrológicos ésta es muy marcada, pues usualmente por la naturaleza de los mismos los datos tienen una variación cíclica.

Para remover los elementos no aleatorios de una serie cronológica se tienen varios criterios, aunque la mayoría involucra el ajuste de un polinomio a los datos, lo cual en una serie hidrológica muchas veces no es factible usar por disponer de muestras pequeñas. Un método alternativo es el de los "Promedios móviles", el cual consiste en encontrar un polinomio que se ajusta a una parte del registro y usar diversos polinomios para las diferentes partes del registro y usar diversos polinomios para las diferentes partes del registro. En los registros hidrológicos es usual considerar, para remover los elementos no aleatorios de una serie de volúmenes de escurrimiento mensual, utilizar para cada mes, el promedio de los volúmenes registrados en dicho mes. Dependiendo del proceso usado para descomponer una serie, se tienen diversos criterios de generación de los mismos, los cuales para analizar una serie pueden combinarse.

8.9.2 Procesos de Generación

Un proceso de generación es la manera por la cual es factible producir una serie cronológica. Algunos procesos pueden expresarse matemáticamente, con lo que es posible determinar directamente las diversas características de la serie cronológica. Si una serie cronológica se aproxima a un cierto proceso, es posible generarla.

El proceso de generación depende como ya se indicó de las características físicas de la serie que se este analizando. En estudios hidrológicos los procesos más usuales son el de los promedios móviles, la suma de armónicas y los autorregresivos.

8.9.2.1 Promedios móviles.

El proceso de los promedios móviles pueden expresarse como

$$x_i = a_0 + a_1 x_{i-1} + a_2 x_{i-2} + \dots + a_m x_{i-m} \qquad (8.52)$$

donde x_i es el valor de la serie en el tiempo t = i, $x_i - 1$ es el valor en el tiempo i − 1, etc. (fig. 8.14) y m son los términos de los promedios móviles. Aquí el problema es determinar cuantos valores de $x_i - k$ influyen en el valor x_i, para posteriormente obtener los coeficientes de peso – – –

$a_o, a_1, a_2 \ldots a_m$. Para ello, se utilizan los coeficientes de correlación seriados r_k entre los diversos valores de x_i y x_{i-k}, el cual como ya se mencionó es análogo al visto en el inciso 8.2, ec 8.19, considerando a los x_i y x_{i-k} como dos grupos de datos. En este caso

$$r_k = \frac{(S\, x_i x_{i-k})}{(S\, x_i x_i \;\; S\, x_{i-k} x_{i-k})^{1/2}} \tag{8.53}$$

donde

$$S\, x_i x_{i-k} = (n-k) \sum_{i=1}^{n-k} x_i x_{i-k} - \sum_{i=1}^{n-k} x_i \sum_{i=1}^{n-k} x_{i-k}$$

$$S\, x_i x_i = (n-k) \sum_{i=1}^{n-k} x_i^2 - \left[\sum_{i=1}^{n-k} x_i \right]^2$$

$$S\, x_{i-k} x_{i-k} = (n-k) \sum_{i=1}^{n-k} x_{i-k}^2 - \left[\sum_{i=1}^{n-k} x_{i-k} \right]^2$$

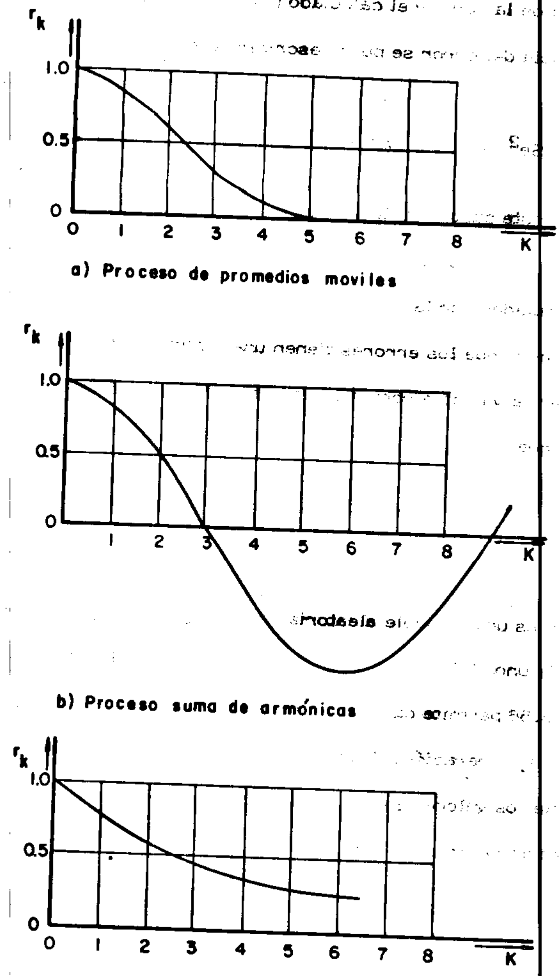
Utilizando la ec 8.53 es posible calcular los coeficientes de correlación seriados r_k. Si la serie sigue un proceso de promedios móviles (ec 8.52) los coeficientes de correlación r_k deberán ser nulos para $k \geqslant m$. Así, calculando los r_k se puede conocer cuál es el proceso que más se apega a la -

serie que se está analizando.

Conocidos los valores de r_k es posible graficar a éstos para los diferentes valores de k, obteniéndose lo que se conoce como correlograma, lo que permite visualizar cual es el proceso más adecuado a utilizar en el análisis de una serie. En la fig. 8.15 se muestran diversos correlogramas según el proceso a que correspondan. Así, en la fig. 8.15 a se tiene el correlograma de un proceso de promedios móviles, en donde se vé que el valor de m corresponde a k = 4, lo que permite acotar la ec 8.52 que define el proceso. No se debe olvidar que este tipo de modelos son teóricos, y que al analizar una serie de datos hidrológicos puede suceder que se tenga una superposición de varios procesos, o bien que debido al tamaño de la muestra, nunca se obtengan coeficientes seriados nulos. Así, al aplicar un proceso como el descrito en la ec 8.52 lo que se hace es considerar los términos de mayor peso, o sea los de coeficientes de correlación seriados altos y despreciar los valores bajos, con lo que al usar el modelo, los valores generados tendrán un cierto error.

De esta manera se tendrá que el proceso de generación se puede escribir como

$$x_i = a_0 + a_1 x_{i-1} + a_2 x_{i-2} + \dots + a_m x_{i-m} - \varepsilon \qquad (8.54)$$



c) Proceso autorregresivo

Fig. 8.15 Correlogramas de diversos procesos donde $\mathcal{E}$ es una variable aleatoria e implica el error que se tiene entre el valor real de la serie y el calculado por la ec 8.52.

La variancia del error se puede escribir de acuerdo con la ec 8.19 como

$$Se^2 = S^2x \ (1 - r^2) \tag{8.55}$$

siendo en este caso Sx^2 la variancia de los valores x_i de la serie de datos en estudio y r el coeficiente de correlación entre los valores de dicha serie y los calculados por la ec 8.52, para el mismo intervalo de tiempo.

Considerando que los errores tienen una distribución normal con media cero al utilizar la variable normalizada y su variancia de acuerdo con la ec 8.55, se tiene que

$$\mathcal{E} = z \ Sx \ (1 - r^2)^{1/2} \tag{8.56}$$

donde z es una variable aleatoria con distribución normal, media cero y variancia uno (Tabla 8.1)

La ec. 8.56 permite calcular el error cuando se utiliza la ec 8.54 como un proceso de generación. Durante el proceso de generación se pueden aceptar constante los valores Sx y r obtenidos de la serie de datos, y valuar el error asignándole valores a por medio de una tabla de distribución normal, considerando la probabilidad de cada valor por medio de una tabla de números -casuales. Si en lugar de generar se desea predecir algun valor, se asigna directamente el valor de la probabilidad con que se desee obtener dicho valor.

Ecuaciones del tipo de la 8.54 son muy usadas para correlacionar dos variables seriadas y en hidrología se utiliza para relacionar los escurrimientos con las lluvias, (capítulo 9). Así por ejemplo, si las x_i son volúmenes de escurrimiento mensual y las y_i son volúmenes de lluvia mensual, para una misma cuenca, se podría plantear un modelo del tipo

$$x_i = a_0 + a_1 y_i + a_2 y_{i-1} + \quad + a_m y_{i-m} + \varepsilon \qquad (8.57)$$

que implica que el volumen del mes i es función del volumen llovido el mismo mes i y los i - k meses anteriores, cuya ecuación es similar a la ec (8.54)

Conocidos los elementos que constituyen la ec 8.52 o la ec 8.54, para obtener los coeficientes de peso $a_0, a_1, a_2 \ldots a_m$ se sigue un proceso similar al visto en el inciso 8.3 al resolver la ec 8.22

8.9.2.2 Proceso de suma de armónicas

En el caso de los coeficientes de correlación seriados r_k de una serie tengan un correlograma como el mostrado en la fig 8.15 b, el proceso de generación de dicha serie sigue una suma de armónicas. En este caso si se deseara generar una serie de datos hidrológicos de este tipo, se tendría primero que quitar la ciclicidad de la misma. Un modelo simple de un proceso de generación de la suma de armónicas es

$$x_i = A \operatorname{sen} \theta\, i + z_i \qquad (8.58)$$

donde A y θ son la amplitud y el período del ciclo respectivamente y una componente aleatoria.

8.9.2.3 Proceso autorregresivo

Procesos de este tipo son muy usados en hidrología pues, con ellos es posible representar los efectos de almacenaje que se tienen en las cuencas. Dentro de los procesos autorregresivos, el más usual es el proceso Markoviano de primer orden, el cual se define como

$$x_i = r_1\, x_{i-1} + \mathcal{E}_i \qquad (8.59)$$

donde r, es el coeficiente seriado de primer orden de las x y $\mathcal{E}$ es una componente aleatoria, e implica que el valor de x_i solo depende del valor en su estado anterior o sea x_{i-1}.

Para un proceso de este tipo, el coeficiente de correlación seriado se puede expresar como $r_k = r_1^k$, cuyo correlograma se muestra en la fig. 8.13c.

Thomas y Fiering* aplicando un proceso markoviano (ec 8.56), generaron escurrimientos mensuales, considerando correlaciones seriadas de éstos. La ecuación de recurrencia utilizada se puede escribir como

$$Q_{i+1} = \overline{Q}_{j+1} + B_j (Q_i - \overline{Q}_j) + S_{j+1} (1 - r_j^2)^{1/2} e_i \qquad (8.60)$$

donde $Q_i + 1$ son los escurrimientos durante el mes i y el mes (i + 1) respectivamente, contados a partir del inciso de la secuencia de generación; $\overline{Q}_j$ y $\overline{Q}_j + 1$ son los escurrimientos medios mensuales durante los meses y (j + 1) respectivamente, dentro de un ciclo anual respectivo de 12 meses; B_j es el coeficiente de correlación para estimar el escurrimiento del mes), del escurrimiento del mes j; $S_j + 1$ es la desviación estandar de los imientos en el mes (j+1); r_j es el coeficiente de correlación entre jos de los meses j y (j + 1); y e_i es una variable aleatoria normal inde te en media cero y variancia uno.

e tipo de ecuaciones también se pueden utilizar para correlacionar por emplo los volúmenes de escurrimientos de dos estaciones de aforo, X y a ec 8.60 se expresa como

* Fiering, M.B.: "Queuing theory and simulation in reservoir design", Trans. Am. Soc. Civil Engrs., vol. 127, pt I, pp. 1114-1144, (1962)

$$Qi_y = \overline{Q}_{jy} + Bj\,(Qi_x - \overline{Q}jx) + S_{jy}\,(1 - r_j^{\,2})^{1/2}\,e_i \qquad (8.61)$$

cuyos valores tienen una interpretación similar a la indicada en la ec. - 8.60