DISTRIBUCIONES DE PEARSON

Pearson desarrolló una familia de distribuciones probabilísticas que pueden ser ajustadas a cualquier distribución impírica.

Función de Densidad

La ecuación general y básica de Pearson es:

$$f(x) = \exp\left[\int_{-\infty}^{x} \frac{a + x}{b_o + b_1 x + b_2 x^2}\, dx\right] \qquad (1)$$

donde a, b_o, b_1, b_2 son constantes.

B. APLICACION A LA FAMILIA DE FUNCIONES DE PEARSON

1.	$K<o$	mejor función:	Tipo I de Pearson sin límite o llamado también distribución Beta
2.	$K<o$	pero K $-\infty$	en este caso varible x positiva y mejor: Tipo III de Pearson
3.	$O<k<1$		Tipo IV de Pearson
4.	$K>1$	:	Tipo VI de Pearson
5.	$K=o$ $\rightarrow$		Distribución normal

VI DISTRIBUCIONES DE PEARSON

A. TIPO I

1. Densidad $f(x) = p_o \left(1 + \frac{x}{a_1} \right)^{m_1} \left(1 - \frac{x}{a_2} \right)^{m_2}$

B. TIPO III ($K = \infty$)

1. Densidad: $p(x) = p_o \left(1 + \frac{x}{a} \right)^{c} e^{-cx/a}$

C. Propiedades

forma de campana o de J o de V

D. Aplicación en Hidrología

1. Bastante uso con computadora (debido a factor de escogencia)

2. No existe papel de probabilidad

3. Existe relación K-T (Chow): factores de frecuencia

4. Pocas ≠ cias con log normal

<u>SELECCION ENTRE DISTRIBUCIONES TEORICAS</u>

I. PARAMETRO DE PEARSON

A. <u>EL PARAMETRO</u>

P

Pearson propuso el uso del parámetro k como criterio para distinguir los principales tipo de funciones de probabilidad o para seleccionar la función o distribución teórica que se ajusta mejor a la distribución empírica:

$$k=\frac{C_s^2\,(E+6)^2}{4(2E-3C_s^2)\,(4E-3C_s^2+12)}$$

donde:

$$C_s=\frac{m_3}{m_2^{3/2}}\approx\frac{a}{s^3}$$

m_3= momento centrado 3er. orden

C_s= coeficiente de sesgo

a = parámetro de sesgo(3er $\approx$ momento)

s = desviación típica.

y

$$a=\frac{n}{(n-1)\,(n-2)}\sum_{i=1}^{N}(x_i-\bar{x})^3$$

Para cálculo:

$$a=\frac{1}{(n-1)\,(n-2)}\left[n\Sigma x_i^3-3\sum_1^n x_i\sum_1^n x_i^2+\frac{2}{n}\left(\sum_1^n x_i\right)^3\right]$$

s = desviación típica = $\sqrt{\text{varianza}}$

$$s=\frac{1}{n-1}\left[\sum_1^n x_i^2-\frac{1}{n}\left(\sum_1^n x_i\right)^2\right]$$

y

$$E=g_2-3$$

donde

$$g_2=\frac{m_4}{m_2^2}$$

m_4= momento centrado de 4° orden

m_2= momento centrado de 2° orden

$$E = \frac{m_4}{m_2^2} - 3$$

$$m_4 = \frac{1}{(n-1)(n-2)(n-3)} \left[(n^2+n)\sum_1^n x_i^4 - 4(n+1)\sum_1^n x_i \sum_1^n x_i^3 - 3(n-1)\left(\sum_1^n x_i^2\right)^2 + 12\left(\sum_1^n x_i\right)^2 \sum_1^n x_i^2 - \frac{6}{n}\left(\sum_1^n x_i\right)^4 \right]$$

Las correcciones se utilizan por estimaciones sesgadas (biased) cuando $n \leq 30$

IV DISTRIBUCIONES EXTREMAS

A. TIPO I (GUMBEL)

1. Densidad $f(x) = \alpha\, e^{-\alpha(x-\beta) - e^{-\alpha(x-\beta)}}$

$$P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx = e^{-e^{-\alpha(x-\beta)}}$$

con

$$y = \alpha(x-\beta) = \frac{x-\beta}{c} \quad \text{variable reducida}$$

$$P(X \leq x) = e^{-e^{-y}}$$

2. Propiedades:

moda : β

(a) media : $\mu = \beta + \frac{0.5772}{\alpha}$

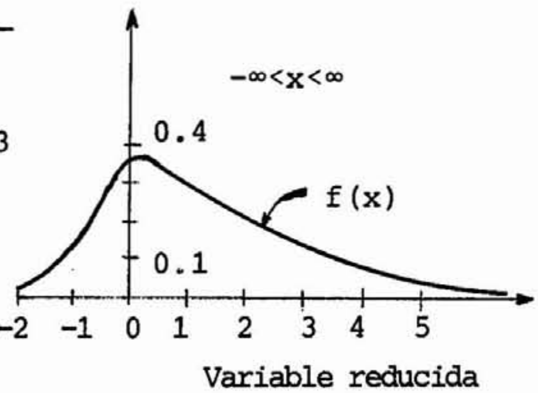
$\mu = x + s \times 0.5772$

(b) desv. típ. $\sigma = \frac{\pi}{\alpha\sqrt{6}}$

(c) Sesgo: $\gamma_1 = 1.29857 \approx 1.3$

(d) curtosis = 4.5

(e) también :



$$\alpha(x-\beta) = -\ln[-\ln P(X\leq x)]$$

de manera que un papel con una doble escala logaritm. para $P(X\leq x)$ y una escala logarit. para x, obtenemos una línea recta. Este gráfico no es comercialmente disponible pero se encuentra en oficinas.

(d) β es la moda : para $x = \beta = m$

$$P(x = \beta) = e^{-1} = 0.368 \quad \text{da moda si } x>o$$
$$y \quad " \quad " \quad = 0.632 \quad " \quad " \text{ si } x<o$$

de modo que en el gráfico β ó m corresponde a la probabilidad 0.368

(e) Si se tiene los dos cuartiles x_1 y x_2 tales que $F(x_1) = 0.25$ y $F(x_2) = 0.75$, se puede hacer una primera aproximación de α, como

$$\frac{1}{\alpha} = 0.6359(x_2-x_1)$$

(f) La media se obtiene para $P(x = \mu) = 0.57$

3. AJUSTE
   (a) gráfico: papel de probabilidades (no comercial sino en oficinas)
   (b) método de momentos y de verosimilitud máxima
   (c) existe relacion K-T.

4. Aplicaciones en Hidrología

   (a) Para valores extemos máximos $\alpha > o$
   (b) Para valores extremos mínimos $\alpha < o$
   (c) Se ajuste bien a los extremos máximos anuales de escorrentía, según Gumbel.
   (d) No se puede justificar uso para sumas de 10 días, ...
   (e) Limitación: no estrictamente aplicable a picos siendo que para pico $x \geq o$

B. TIPO II (Weibull)

1. densidad: $$f(x) = - \frac{\alpha}{\beta - \gamma} y^{(\alpha-1)/\alpha} e^{-y}$$

a. Para extremos mínimos

$$P(x \leq x) = \exp\left[-\left(\frac{x-\gamma}{\beta-\gamma}\right)^{\alpha}\right]$$

con $$y = \left(\frac{x-\gamma}{\beta-\gamma}\right)^{\alpha} \quad y \quad xz\gamma \;, \quad \alpha > o \;, \quad \beta > \gamma$$

b. Para extremos máximos $\quad y = \dfrac{\gamma - x}{\gamma - \beta}$

$$P(X \leq x) = e^{-y} \qquad \text{donde } x \leq \gamma$$

$\alpha > o$ y $\beta < \gamma$

2. Propiedades

(a) Para $x = \beta$ (moda) $P(X=\beta) = 0.368$

(b) $\ln\left[-\ln P(X \leq x)\right] = \alpha \ln(x-\gamma) - \alpha \ln(\beta - \gamma)$

$\therefore$

Sólo $(x - \gamma)$ con $P(X \leq x)$ da 1 línea recta en papel long. y doble-log escalas. A menos de que $\gamma = o$ entonces x con P da recta.

C. AJUSTE

(a) Gumbel mostró que se ajuste bien a extremos mínimos anuales

(b) Para T= 2.33 o $P(X \leq \bar{x}) = 42.9\%$ se obtiene la $\bar{x}$.

V DISTRIBUCION BINOMIAL (discreta)

$$p(x) = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}$$

$q = 1 - p.$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

RIESGO

La probabilidad de que la avenida de probabilidad p, ocurra por lo menos una vez durante r años $= 1 - (1 - p)^r$

Ejemplo:

Si $T_r = 10$ años $\qquad p = 0.10$

$N = 10$ años $= T = n$

$$P(\text{ no ocurrencia}) = \frac{T_r!}{(T_r - o)!\,o!}\, p^o (1-p)^{T_r}$$

$$= (1-p)^r$$

$$\text{Riesgo} = P(\text{ por lo menos 1 vez}) = 1 - (1-p)^r$$

$$\text{Riesgo} = 1 - (1-0.10)^{10} = 1 - (0.90)^{10}$$

EJEMPLO 2 :

(a) Cuál es la probabilidad de que una creciente mayor a la creciente de período de retorno de una vez en 5 años en el promedio, no ocurra en un período de 10 años?

(b) Cuál es las probabilidades de que 1, 2, 3, y 4 de estas crecientes ocurrían?

Solución

(a) $X = o$ ocurrencia $\qquad p = \frac{1}{T_r} = \frac{1}{5} = 0.20$

$n = 10$

$$P(X) = \frac{10!}{(10-0)!\,0!} (0.20)^0 (0.80)^{10} = 0.11$$

(b) $P(X=1) = \frac{10}{9!\,1!} (0.20)^1 (0.80)^9 = 0.27$

$P(X=2) = 0.28$

$P(X=3) = 0.20$

$p(X=4) = 0.11$

GENERALIZACION

r años generalmente representa la vida útil de la obra, p = la probabilidad de diseño a utilizar para un riesgo igual a

$[1 - (1-p)^r]$

Como

$p = \frac{1}{T_r}$ podemos decir que para este riesgo habría que utilizar para el diseño un período de retorno T_r (generalmente $\neq$ r)

Ilustración

Se ha de construir una presa durante un período de 3 años. Para mantener seco el tramo de construcción, se construyó temporalmente diques y obras de derivación para resistir la avenida de 8 años.

Determinar el riesgo de que los diques sean inundados:

1) por lo menos una vez durante el período de construcción.
2) Exactamente 2 veces durante el período de construcción
3) Solamente el 3er año durante el período de construcción

Solución:

1) $P(x>1) = 1 - P(x=o) = 1 - \binom{3}{o} p^o (1-p)^3$

como $p= \frac{1}{T_p} = \frac{1}{8} = 0.125$

$$P(\text{por lo menos una vez}) = 1 - \frac{3}{3!\,0!}(1-0.125)^3 = 1-0.67$$

$$= 0.33$$

Existe 33% chance de que los diques sean inundados por lo menos una vez durante los 3 años de construcción.

2) $P(x=2) = \binom{3}{2} p^2 (1-p) = \frac{3!}{2!\,1!} (0.125)^2 (1-0.125)$

$P(x=2) = 0.041$

Hay una probabilidad de 4.1% de que la avenida de diseño sea sobrepasado 2 veces.

3) $P\ (\text{sólo el 3er. año}) = (1-p)(1-p)p = (1-p)^2$

$$p = (1-0.125)^2 (0.125)$$

$P\ (\text{sólo en 3er. año}) = 9.57\%$

Es también igual a la probabilidad de ocurrencia 1 sola vez, dividida por el número de permutación de 1 vez en 3 años ó

$$P(\text{sólo el 3er. año}) = \frac{P(x=1)}{3} = \frac{1}{3}\left(\frac{3!}{1!\,2!}\right) (0.125)^1$$

$$(1-0.125)^2 = 9.57\%$$

I. AJUSTE GRAFICO

A. Papel de probabilidad

1. Normal (línea recta) puesto que (s=o)
2. log. normal (no necesarimanete línea recta)

3. extremo tipo I [escalas: normal para x vs ln(ln F/x)]

(línea recta puesto que C_s=1.3 const.)

4. Extremo tipo II (escalas [ln for (x-β) vs ln(ln F(x))]

(no necesariamente una recta)

III. ANALISIS DE FRECUENCIA MEDIANTE FACTORES DE FRECUENCIA

(Chow)

A. ECUACION GENERAL DEL ANALISIS DE FRECUENCIA

1. La variable aleatoria hidrológica puede escribirse como:

$$x = \bar{x} + \Delta x \qquad (1)$$

donde

Δx la desviación depende del grado de dispersión: función de σ

$$\Delta x = K\,\sigma \qquad (2)$$

∴ donde

K = factor de frecuencia

Combinando (1) y (2)

$$x = \bar{x} + \sigma K \qquad (3)$$

ó

$$\frac{x}{\bar{x}} = 1 + \frac{\sigma}{\bar{x}} K \qquad (4)$$

Reconociendo $\frac{\sigma}{\bar{x}} = c_v$ = coeficiente de variación por definición.

$$\frac{x}{\bar{x}} = 1 + c_v K \qquad (5)$$

- Ec. (5) fue propuesta por Chow como la ecuación general del análisis de frecuencia.
- Ec. (5) es aplicable a varias distribuciones teóricas.
- Para 1 dist. teórica una relación entre K y T (o P) puede ser desarrollada, llamada las curvas K-T, presentada en forma de curvas o tablas.

  donde T = intervalo de recurrencia ó período de retorno.

2. Curvas K-T

(a) Aplicación: calcular parámetros $\bar{x}$ C_v

- para cada valor de x, el factor de frecuencia K puede ser determinado mediante la ecuación (3).
- con el valor de K obtenido de (3) se entra en las curvas K-T para determinar el valor de T correspondiente a cada x, obteniendo así la distribución teórica se obtiene.

(b) Empirical K-T

Fuller derivó empíricamente para caudal diario máximo anual

$$x = \bar{x}\ (1 + 0.8 \log T) \qquad (6)$$

comparando con (5)

$$K = \frac{0.8}{c_v} \log T$$

c_v varía de 0.1 a 2.0 con un valor promedio de 0.50. Sin embargo la fórmula de Fuller es empírica y desde Fuller otros métodos más analíticos han sido empleados.

(c) Distribución Normal

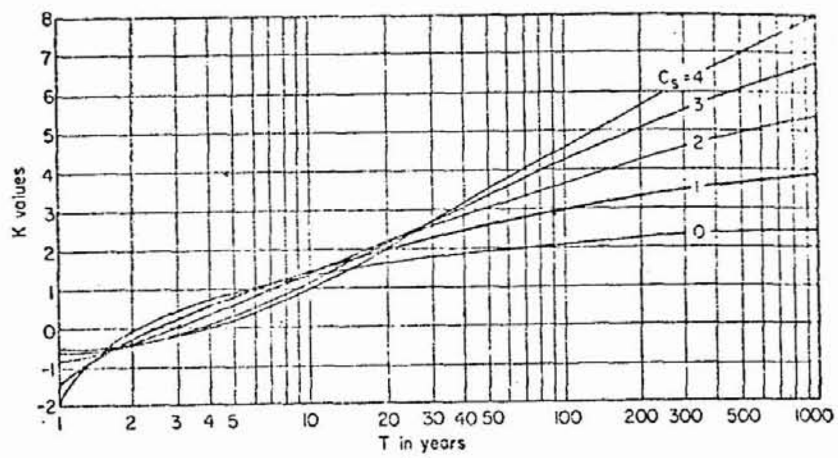
$$\therefore \quad K = \frac{x - \bar{x}}{\sigma}$$

$$P\ (x < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K} e^{-\frac{K^2}{2}}\ dK$$

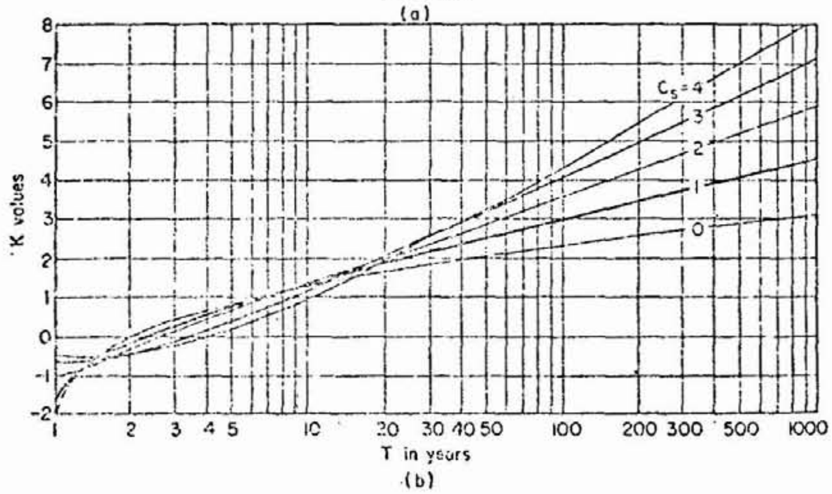
$$T = \frac{1}{P(x \geqslant x)} = \frac{1}{1-P(x \leqslant x)}$$

Valores de $P(x \leq x)$ ó T para varios valores de K pueden ser hallados de una tabla de probabilidad normal

(d) Distribución de Pearson (según Foster)



(a)



(b)

donde $c_s = \frac{\alpha}{\sigma^3} \approx \frac{a}{s^3}$ (1)

y $$a = \frac{N}{(N-1)(N-2)} \sum_{i=1}^{N} (x_i - \bar{x})^3$$ con $$\alpha = \frac{1}{N} \sum_{i=1}^{N} (x_i - Z)^3$$

$$= \frac{N^2}{(N-1)(N-2)} (\bar{x}^3 - 3\bar{x}^2 \bar{x} + 2\bar{x}^3)$$

c_s = coeficiente de sesgo

a = parámetro de sesgo

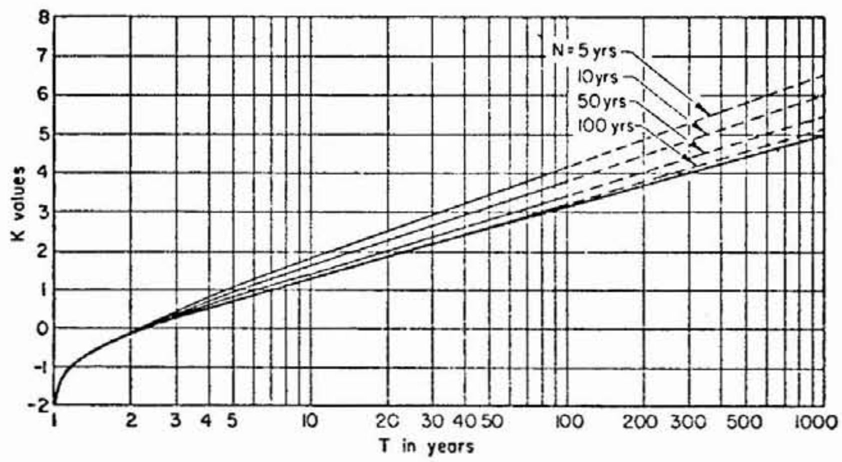
s = desviación típica

Foster recomendó multiplicar c_s de (1) por (1 + 8.5/N) y por 1 + 6/N para tipo III.

(e) Distribución extremal. Tipo I
(distribución de tipo Gumbel)
(de coeficiente de sesgo constante c_s = 1.139)

Chow mostró que:

$$K = -\frac{\sqrt{6}}{\pi} \left[\gamma + \ln \ln \left(\frac{T}{T-1}\right) \right]$$



donde

= 0.57721 Constante de Euler

Cuando $x = \bar{x}$ la ecuación (1) da K = o.

y $\therefore$ T = 2.33

recurrencia de la creciente media anual o de la media de la distribución extremal tipo I.

La distribución extremal Tipo I ó Gumbel Tipo I tiene el coeficiente de sesgo constante $C_s = 1.139$

(1) puede escribirse:

$$T = \frac{-\exp\{\exp\left[-(\frac{\pi}{\sqrt{6}} K + \gamma)\right]\}}{1 - \exp\{\exp\left[-\frac{\pi}{\sqrt{6}} K + \gamma)\right]\}}$$

(f) Distribución lognormal

Chow:

$$K = \frac{e^{\sigma_Y K_Y - \sigma_Y^2/2} - 1}{(e^{\sigma_Y^2} - 1)^{1/2}}$$

donde:

$$x = e^Y \qquad y \qquad K_Y = \frac{Y - \bar{Y}}{\sigma_Y}$$

Chow presenta una tabla dando K para diversos valores de $P(X \leq n)$ para diferentes valores de c_s.

RELACION K-T PARA DISTRIBUCION EXTREMAL TIPO I

Chow derivó la relación K-T siguiente:

$$K = -\frac{\sqrt{6}}{\pi}\left[\gamma + \ln\ln\left(\frac{T}{T-1}\right)\right]$$

$$-\frac{\pi}{\sqrt{6}}K - \gamma = \ln\ln\left(\frac{I}{T-1}\right)$$

$$\frac{T}{T-1} = \exp\{\exp\left[-\left(\frac{\pi}{\sqrt{6}}K + \gamma\right)\right]\}$$

$$T = (T-1)\exp\{\exp\left[-\left(\frac{\pi}{\sqrt{6}}K + \gamma\right)\right]\}$$

$$T = \frac{-\exp\{\exp\left[-\left(\frac{\pi}{\sqrt{6}}K + \gamma\right)\right]\}}{1-\exp\{\exp\left[-\left(\frac{\pi}{\sqrt{6}}K+\gamma\right)\right]\}}$$

Ej.

Supongamos que $K = 1.77$ y $\gamma = 2 \rightarrow T = 1.39$

Para Gumbel

Ilustración

La media de los caudales máximos anuales para una estación de 25 años de registro, es 1000 m^3/s. La desviación típica es 400 m^3/s. Estimar la magnitud de la crecida de 50 años de período de retorno.

Solución

Para T = 50 la Fig. 8-1-7 (curva para distribución extremal tipo I),

$K = 3.088$

$\therefore$

$$x = \bar{x} + Ks$$

$$x = 1000 + 3.088\ (400) = 2235\ \text{cfs}$$

III. METODOS ANALITICOS DE AJUSTE

A. Principio

1. El método analítico consiste en estimar los parámetros poblacionales desconocidos de la distribución teórica a partir de los parámetros muestreales.

2. El segundo paso es de utilizar una prueba de bondad de ajuste para determinar si el ajuste de la distribución teórica a la distribución empírica es aceptable.

3. De modo que los métodos analíticos de ajuste se identifican a los métodos de estimación de parámetros poblacionales.

B. Métodos

1. Método gráfico (papel de probabilidad y su uso en la determinación de parámetros).

2. Métodos de mínimos cuadrados

Siendo conocida la función de distribución teórica $f(x, \alpha, \beta)$ la suma de los cuadrados de las diferencias es

$$S = \sum_{1}^{N} (y_i - y)^2 = \sum_{1}^{N} \left[y_i - f(x_i,\alpha,\beta,\gamma,..)\right]^2$$

Estimar α,β,γ ... por a, b, c

Para mínimo

$$\frac{\delta S}{\delta a} = 0 \qquad \frac{\delta S}{\delta b} = 0 \qquad \cdots\cdots$$

Así podemos calcular a, b, c, ... , que son estimaciones muestreales de los parámetros poblacionales α, β, γ . . .

3. Método de los Momentos

   Basado en las relaciones entre momentos y parámetros

4. Método de la verosimilitud máxima.

   Definir: $$L = \prod_{i=1}^{N} f(x_i, \alpha, \beta, \gamma \ldots)$$

   $$\ln L = \sum_{i=1}^{N} \ln f(x_i;\, a, b, c \ldots)$$

   Para L máx. tenemos:

   $$\frac{\delta \ln L}{\delta a} = 0 \; ; \qquad \frac{\delta \ln L}{\delta b} = 0 \; ; \quad \ldots$$

ILUSTRACION DE AJUSTE

I. AJUSTE ANALITICO

A. Distribución extrema Tipo I

1. Estimación de parámetros

Por el método de los momentos obtenemos:

$$\beta = \bar{x} - 0.450\ s$$

$$\alpha = \frac{1.281}{s}$$

2. Datos

Año Hidrol.	Precip. máxima en 24 h. (mm)	Orden	X	Probabilidad distr.empír. % $\frac{m}{n+1}$	Período de retorno T_r (años)
1937-38	137	1	137	6.25	16
39-38	82	2	127	12.25	8
39-40	58	3	82	18.75	5.33
40-41	58	4	74	25.00	4.00
41-42	127	5	64	31.25	3.20
42-43	74	6	58	37.50	2.67
43-44	64	7	58	43.75	2.29
44-45	55	8	55	50.00	2.00
45-46	47	9	51	56.25	1.78
46-47	51	10	47	62.50	1.60
47-48	38	11	41	68.75	1.45
48-49	33	12	40	75.00	1.33
49-50	41	13	38	.81.25	1.23
50-51	30	14	33	87.50	1.14
51-52	40	15	30	93.75	1.07

$\Sigma x_i = 935$ $\bar{x} = 62.33$

$\Sigma x_i^2 = 72471$ $s = 30.76$

$$\therefore\ \beta = \bar{x} - 0.450\ s = 62.33 - 0.450(30.76) = 48.49$$

$$\alpha = \frac{1.281}{30.76} = 0.0416$$

$\therefore$

$$F(x) = e^{-e^{-0.0416(x-48.49)}}$$

Luego para cada x observado calcular F(x) correspondiente y aplicar prueba de Kolmogorov o de x^2.

PRUEBA DE BONDAD DE AJUSTE

PRINCIPIO: Tal prueba requiere un parámetro para medir el ajuste. Discutiremos 2 parámetros:

(a) chi cuadrado

(b) Smirnov-Kolmogorov

I. PRUEBA DE SMIRNOV-KOLMOGOROV

1. Estadística de Smirnov-Kolmogorov

a. Sea una muestra (registro) ordenada en orden ascendente como:

$$x_1, x_2, x_3, \quad . \ . \ . \quad , x_n$$

u ordenada en orden descendente como:

$$x_1', x_2', x_3', \quad . \ . \ . \quad , x_n'$$

b. Calculamos las probabilidades $P(x_i)$ como

$$P(x_i') = \frac{m}{n + 1}$$

c. Plotear la distribución empírica $P(x_i')$ vs. x_i.

d. Calcular la estadística de Smirnov-Kolmogorov definida como:

$$\Delta = \max\left[F(x) - P(x')\right] \quad (1)$$

Si Δ_o es el valor crítico de Δ para el nivel α de significación definido como:

$$P\ \{ \max[F(x) - P(x')] \geqslant A_o \} = \alpha$$

e. tiene una distribución de la forma:

$$\phi\ (z) = P\ (\ \sqrt{n}\ \Delta_o) \leqslant z$$

$$\lim_{n \to o} \phi\ (z) = K(z) = \sum_{-\infty}^{\infty} (-1)^K\ e^{-2K^2 . z^2}$$

f. Esta distribución está tabulada (Yedjevich pp.229) valores de Δ_o para valores de n y de α.

2. Procedimiento de la prueba

a. Obtener la desviación máxima entre la probabilidad de la distribución empírica y la de la distribución ajustada de la manera siguiente: $\Delta = [F(x) - P(x_i)]$

(i) para ajuste gráfico a una distribución teórica o a una línea cualquiera (no una distribución teórica), obtener la desviación máxima por medición directa.

(ii) para ajuste analítico o mediante factor de frecuencia, calcular para cada punto F(x) (distribución teórica) y $P(x_i)$ (distribución empírica), hacer las diferencias y obtener la diferencia máxima.

b. De la tabla 10.3 (pp.229 Yevjevich) obtenemos para n y para un nivel α escogido a priori el valor de Δ_o.

c. Si $\Delta < \Delta_o$ el ajuste es aceptable por la prueba de Smirnov-Kolmogorov.

3. Ventajas y Limitaciones

a. Ventajas

(i) no requiere un conocimiento a priori de F(x) que puede cualquier distribución no teórica.

(ii) comparándola con la prueba x^2 no hay condición de que cada clase de frecuencia debe contener más de 5 miembros o valores.

(iii) no se requiere hacer intervalos de clase.

b. Desventajas

(i) no es una prueba exacta sino una prueba aproximada.

II. LA PRUEBA X^2

A. El estadístico X^2 (muestral)

1. Definición:

$$X^2 = \frac{(O_1 - e_1)}{e_1} + \frac{(O_2 - e_2)}{e_2} + \ldots + \frac{(O_K - e_K)^2}{e_K}$$

ó

$$X^2 = \sum_{i=1}^{K} \frac{(O_i - e_i)^2}{e_i}$$

con

$$\sum_{i=1}^{K} O_i = \sum_{i=1}^{K} e_i = n$$

k = número de intervalos de clase

n = número de datos

2. Distribución

$$y = y_o\, X^{\nu-2}\, e^{-1/2\, X^2}$$

donde:

y → ordenada la curva de densidad

y_o → constante par a limita el valor de y a 1 máx. de 1

ν → número de grados de libertad

3. Grados de libertad ν

(a) $\nu = k-1$

si las e_i pueden calcularse sin estimar estadísticas poblacionales con estadísticas muestreales, por sí son conocidas (k-1) frecuencias, la frecuencia restante puede ser determinada mediante

$$\sum_{i=1}^{k} p_i = 1$$

(b) $\nu = k-1-m$

si las e_i se calculan estimando m parámetros de la población a partir de las estadísticas muestreales.

B. Ensayo de significación (sólo para distribución de datos agrupados)

1. Procedimiento

(a) Escoger el nivel de significación α (en general $\alpha = 0.05$).

(b) calcular $X^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}$

(c) si $X^2 = 0$ ó si $X^2 < X^2_{(1-\alpha)}$ ajuste aceptable si $X^2 > X^2_{(1-\alpha)}$ ajuste malo.

ANALISIS REGIONAL DE FRECUENCIA

Método

(1) Se seleccionan dentro de la región las estaciones con más de 5 años de registro; se tabulan los datos de crecientes anuales máximos, cuando el área contribuyente de dos estaciones sobre un mismo río no difiere en más del 25% se puede considerar como una sola estación.

(2) Se elabora un diagrama de barras con la longitud de registro para cada estación. Se toma el período base generalmente como aquel que posee mayor número de años de registro. Se estiman los datos faltantes, generalmente usando correlaciones, de las crecientes máximas anuales de un río contra las de otro (no necesariamente para la misma fecha de ocurrencia).

(3) Para cada estación, usando papel Gumbel-log ó Gumbel aritmético se traza la línea de frecuencia con $Tr = \frac{n+1}{m}$ (los datos interpolados no se usan en la curva de frecuencia directamente, pero sirven para dar el valor correcto a m para calcular el Tr de las crecidas registradas.

(4) A partir de (3) se obtienen los valores de $Q_{2,33}$, que por definición es la creciente media anual. Igualmente se obtiene Q_{10}, ésto es la crecida correspondiente a Tr = 10 años.

(5) Prueba de Homogeneidad: (Tabla 3)

a. Hallar $\frac{Q_{10}}{Q_{2,33}}$ para cada estación

b. Hallar el promedio de los valores $\frac{Q_{10}}{Q_{2,33}}$ =

$$= \Sigma \frac{\frac{Q_{10}}{Q_{2,33}}}{n^\circ \text{ estaciones}}$$

c. Se obtiene para cada estación los valores de:

$$Q_{2,33} \times \left(\frac{\frac{Q_{10}}{Q_{2,33}}}{N^\circ \text{ estac.}}\right)$$ (valor teórico medio de la crecida de 10 años).

d. Al valor teórico medio de la crecida de 10 años se le calcula a partir de las curvas de frecuencia el tiempo de retorno correspondiente T.

e. Se calcula el tiempo de registro ajustado (TA) como el correspondiente al de las observaciones más la mitad del registro obtenido por correlación.

f. Se llevan estos valores a un gráfico con las curvas de prueba, y se usan solamente los registros correspondientes a las estaciones que se ploteen dentro de las curvas.

Tabla 3

Valores para la prueba de homogeneidad

N°	Río	Area	$Q_{2,33}$	Q_{10}	$\frac{Q_{1o}}{Q_{2,33}}$	$DxQ_{2,33}$	T correspond. a la colum.	Período ajustado de reg.
1	2	3	4	5	6	7	7	8

$$\Sigma \left(\frac{\frac{Q_{10}}{Q_{2,33}}}{N^\circ \text{ estaciones}} \right) = D$$

(6) Para cada estación se determinan las relaciones de crecidas $Q1,1/Q_{2,33}$, $Q1,5/Q_{2,33}$ etc., en una tabla igual a la Tabla 4. Se calculan las medianas de estas relaciones de crecidas:

Medianas: el valor central del conjunto ordenado cuando el número de observaciones es impar o el promedio de los dos valores centrales, cuando el número de observaciones es par.

Tabla 4

Mediana de las Relaciones de Crecidas $\frac{Q}{Q_{2,33}}$

Estación	Tiempo de Retorno en Años					
	1.1	1.5	5	10	20	50

(7) Definición de la curva regional de frecuencias. En papel de Gumbel se grafican las medianas de las relaciones de crecidas contra el período de retorno correspondiente. Esta curva que muestra la relación crecida-crecida anual media contra tiempo de retorno se basa en todos los registros existentes para la región, y es representativa de toda el área (Fig. 1)

(8) Se hace en papel de log-log una correlación entre crecida media anual y área de captación (Figura 2).

(9) Para determinar la frecuencia de crecidas de cualquier sitio se determina:

a. $Q_{2,33}$ a partir del área de drenaje de la cuenca.

b. $Q/Q_{2,33}$ para un período de retorno dado, a partir de la curva $Q/Q_{2,33}$ vs. Tr.

c. Se multiplican los dos valores.

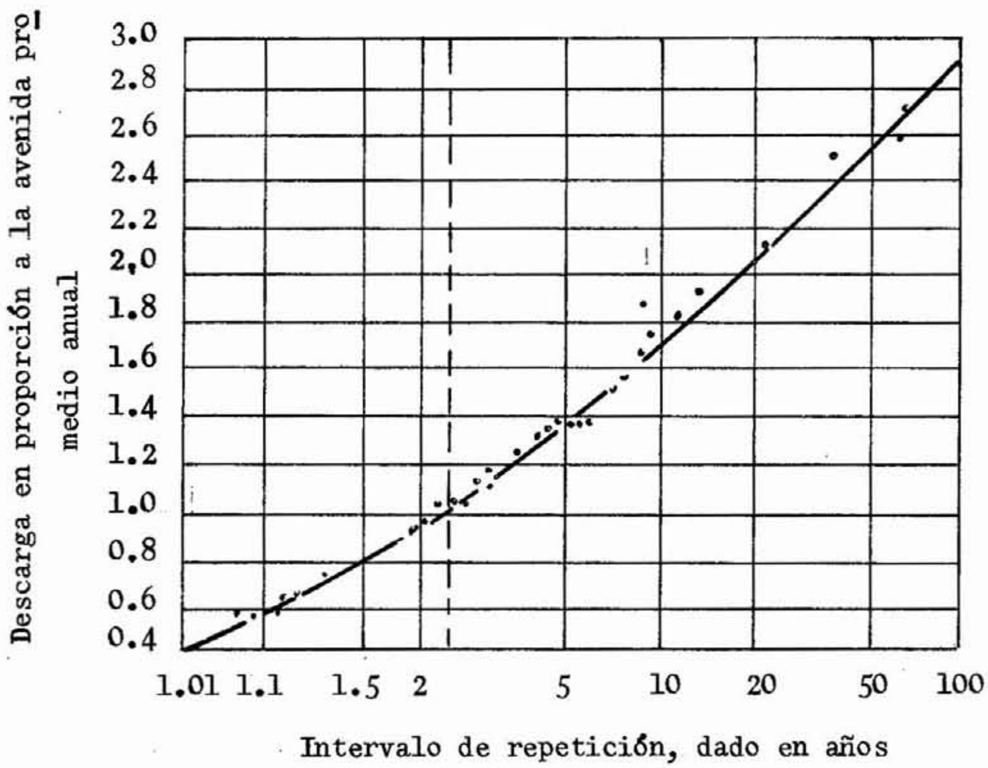


Fig. 1. Curva de Frecuencia Regional

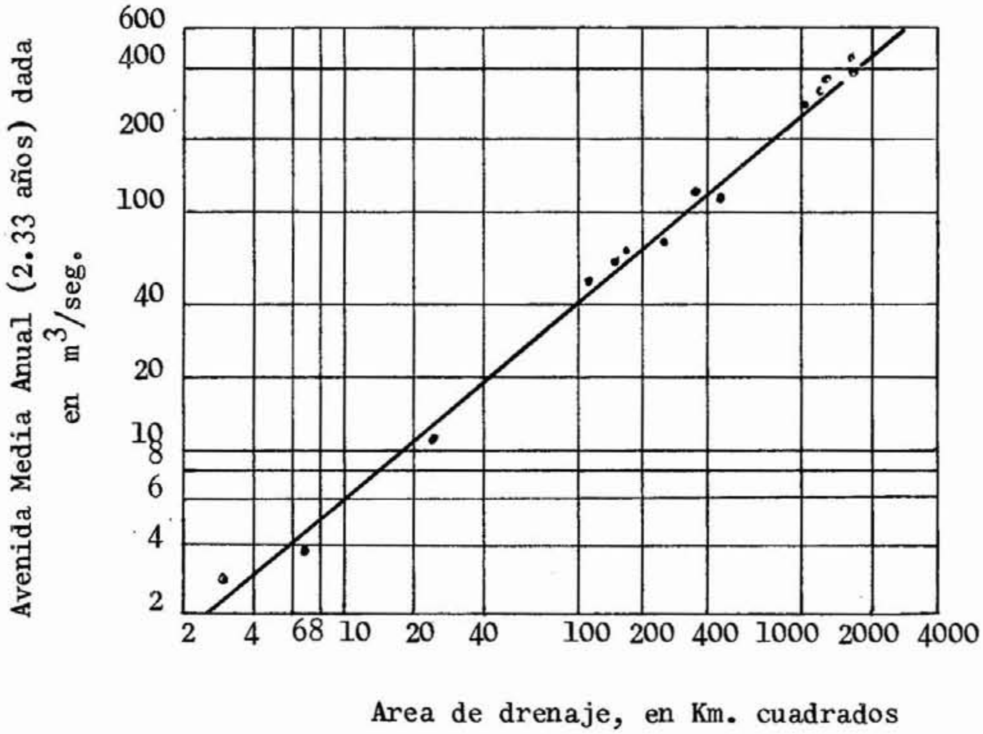
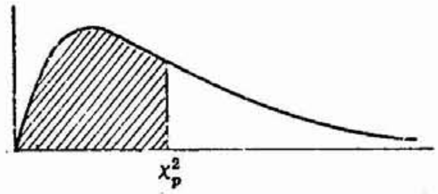


Fig. 2. Variación de la Avenida Media Anual con el Area de Drenaje

Apéndice IV

PERCENTILES (χ^2_p) DE LA DISTRIBUCION CHI-CUADRADO CON ν GRADOS DE LIBERTAD (AREA SOMBREADA = p)



ν	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$	$\chi^2_{0,75}$	$\chi^2_{0,50}$	$\chi^2_{0,25}$	$\chi^2_{0,10}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	7,88	6,63	5,02	3,84	2,71	1,32	0,455	0.102	0,0158	0,0039	0.0010	0.0002	0,0000
2	10,6	9.21	7,38	5,99	4,61	2,77	1.39	0,575	0.211	0,103	0,0506	0.0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0.216	0,115	0.072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7.34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33.4	30,2	27.6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37.2	34,8	31,5	28,9	26,0	21,6	17.3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23.8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36.8	33,9	30,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20.8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40.1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	33,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29.3	24,5	20,6	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51.8	45.6	39,3	33,7	29,1	26,5	24.4	22,2	20.7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	42,9	37,7	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	52,3	46,5	43,2	40,5	37,5	35,5
70	104.2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51.7	48,8	45,4	43,3
80	166,3	112,3	106.6	101,9	96,6	88.1	79,3	71,1	64,3	60,4	57.2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3

Procedencia: Catherine M. Thompson. *Table of percentage points of the χ^2 distribution*, Biometrika, Vol. 32 (1941), con permiso de los autores y editores.

PROBABILIDAD Y ESTADISTICA EN HIDROLOGIA

I. AMBIENTES, SISTEMAS, PROCESOS Y MODELOS HIDROLOGICOS

A. AMBIENTES HIDROLOGICOS:

1. Definición: Son aquéllas partes del universo y de la tierra en las cuales ocurren o toman lugar fenómenos hidrológicos particulares.

2. Ejemplos:

   a) océanos y mares como abastecedores principales de agua al ciclo hidrológico mediante evaporación.

   b) La atmósfera o el aire como: 1) almacenamiento de las aguas evaporadas 2) vehículo de transporte de las aguas" entre ≠ puntos de la tierra 3) ambiente para la condensación y la precipitación.

   c) Superficie de suelo y agua como almacenaje, recipiente de la precipitación y abastecedores de agua a la atmósfera y al suelo mediante la evaporación y la infiltración.

   d) La parte superior de la corteza terrestre como almacenamiento de humedad del suelo y de las aguas subterráneas.

   e) Las plantas como almacenamiento de intercepción y abastecedor de agua por transpiración a la atmósfera.

3. Consecuencias:

- Estos diferentes ambientes afectan a su manera particular a la calidad del agua.
- La cantidad de agua a encontrarse en cada ambiente varía de acuerdo al ambiente y asimismo la distribución cuantitativo del agua es función de los ambientes bajo consideración.
- Cada ambiente puede ser considerado como un sistema con propiedades particulares.
- En la mayoría de los problemas hidrológicos se supone que el ambiente no cambia en el tiempo, o mejor, que sus variaciones en el tiempo son neglegibles cuando se las compara con los procesos hidrológicos. Una excepción lo constituye el proceso del transporte y depósito de sedimento que puede modificar rápidamente el ambiente.

B. PROCESOS HIDROLOGICOS

1. Definición de un Proceso

A cualquier fenómeno sometido a cambios contínuos (particularmente con respecto al tiempo) se le llama un proceso.

2. Definición de un Proceso Hidrológico

a) Cualquier ambiente hidrológico recibe agua, la almacena y la deja salir hacia un otro ambiente.

b) Se puede conceptualizar estos fenómenos que toman lugar en un ambiente hidrológico como entrada, almacenamiento y salida.

c) Entrada, almacenamiento y salida son fenómenos sometidos a cambios contínuos tanto en el tiempo como en el espacio, pues representan cada uno un proceso hidrológico. En general, hay más de una variable involucrada en 1 proceso.

d) Las entradas y salidas al ambiente las designamos como procesos de transferencia y el proceso de almacenamiento en el ambiente lo llamamos un almacemiento.

e) Mediante los procesos de transferencia el agua es transferida de un almacenamiento a otro, de un punto o región de un almacenamiento a otro punto o región del mismo almacenamiento.

f) Ejemplos de procesos hidrológicos:

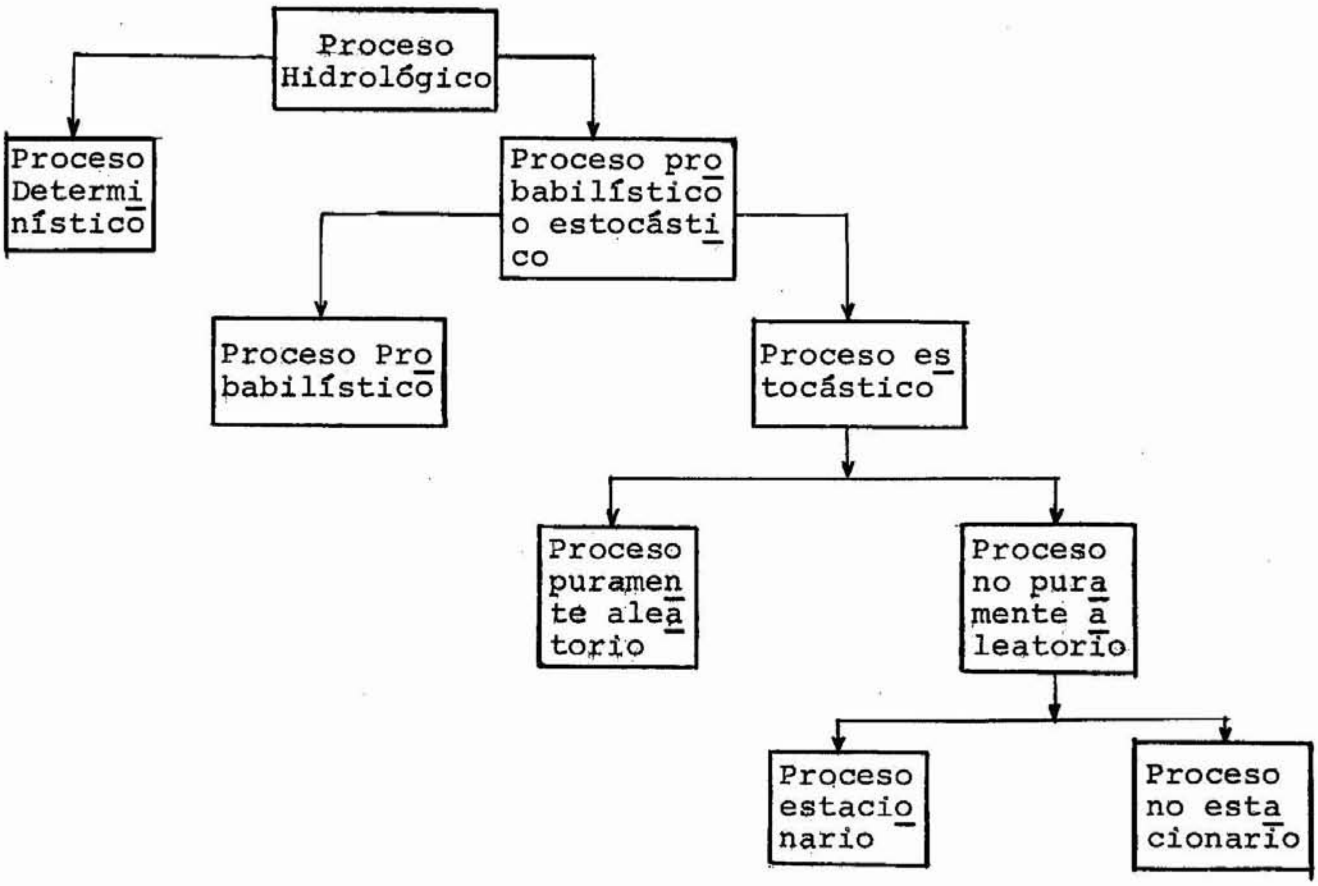
   (i) la precipitación: proceso mediante el cual el agua es transferida de la atmósfera a la superficie de la tierra.

   (ii) el escurrimiento mediante el cual el agua pasa del almacenamiento (momentáneo) de detención en la superficie del suelo a otro punto de este almacenamiento o al almacenamiento de detención en el cauce de un río.

g) El almacenamiento no es idéntico al ambiente hidrológico que la contiene, puesto que en general no ocupa toda la capacidad de almacenaje del ambiente e involucra cambios contínuos en el tiempo, mientras que el ambiente no cambia tan rápidamente.

h) Ejemplos de almacenamiento:
   (i) El agua almacenada en un embalse medio lleno.
   (ii) el almacenamiento de humedad del suelo y el almacenamiento de agua subterránea.

3. CARACTERISTICAS DE LOS PROCESOS HIDROLOGICOS.



a. PROCESOS DETERMINISTICOS

(i) Cuando se ignora la probabilidad o chance de ocurrencia de las variables involucradas en un proceso, se considera que este último obedece a una ley definida de certidumbre y no a ninguna ley probabilística, entonces se describe el proceso como determinístico.

(ii) Procesos determinísticos resultan particu larmente de las leyes de la mecánica de los fluidos y de la termodinámica.

(iii) Procesos hidrológicos puramente determinísticos pueden ser realizados solamente bajo condiciones controladas y artificiales.

(iv) Ejemplos de Procesos Hidrológicos Determinísticos:

(α) El escorrimiento en una superficie impermeable, causado por una entrada bien especificada de lluvia, con pérdidas neglegibles por evaporación e infiltración. Bajo condiciones naturales intervendrían muchos factores aleatorios.

(β) La curva de calibración de una relación unívoca entre caudal y nivel en una sección transversal con un lecho fijo y estable completamente. Para lecho movible, factores aleatorios influyen en la relación o curva.

b. PROCESOS PROBABILISTICOS O ESTADISTICOS.

(i) Cuando se toma en cuenta el chance o probabilidad de ocurrencia de los valores de las variables de proceso hidrológico, entonces hablamos de procesos probabilísticos o estocásticos. En otros términos las variables obedecen a las leyes de probabilidad.

(ii) Procesos Probabilísticos

α) Definición. En un proceso probabilístico se ignora la secuencia de ocurrencia cronológica de las variables. Se asume que las variables siguen una distribución probabilística en la cual las variables se consideran puramente aleatorias. De modo que un proceso probabilístico es completamente independiente del tiempo.

β) Ejemplos de Procesos Probabilísticos:

Precipitación máxima anual o la escorrentía máxima anual, cuando se ignora la secuencia de ocurrencia de estos valores.

(iii) Procesos Estocásticos.

α) Definición. Cuando se toma en cuenta la secuencia de ocurrencia cronológica de las variables y su probabilidad de ocurrencia, hablamos de un proceso estocástico. En un proceso estocástico las variables pueden ser tratadas como puramente aleatorias o como no puramente aleatorias.

β) Aleatorios Puros.
En este caso los miembros de la serie cronológica son independientes entre sí y constituyen una secuencia aleatoria.

γ) Aleatorios No Puros.
Los miembros de la serie son independientes el uno del otro y son compuestos de un componente determinístico y de un componente aleatorio puro, asimismo consti-

tuyendo una secuencia no aleatoria pura.

ρ) Estacionarios

1) Definición. Si la distribución probabilística de una variable de un proceso estocástico queda constante durante o a través del proceso, entonces el proceso y la serie cronológica son estacionarios.

2) Ejemplos.
El registro de los valores vírgenes del caudal de un río sin cambios significantes en las condiciones características o climáticas durante el período del registro, se consideran como una serie cronológica estacionaria.

θ) No Estacionarios.

1) Definición. Cualquier proceso estocástico que no sea estacionario. Estacionaridad de valores de caudales puede resultar de la intervención del hombre, de la inconsistencia y no homogeneidad de los datos.

2) Ejemplos

Una serie no estacionaria puede ser constituída por un registro de 20 años de caudal, con 5 años de registros obtenidos después de construir una presa aguas arriba de la estación fluviamétrica.

Registros de caudal afectados por modificaciones naturales y lentas de las condiciones de la lluvia o de la escorrentía.
En estos ejemplos puede ser que cambia la media de la distribución o su varianza , de modo la distribución no queda constante.

3) Los procesos hidrológicos son esencialmente estocásticos o una combinación de procesos determinísticos y estocásticos.

ESTADISTICA

I. Definición:

Bajo este término se agrupan los métodos científicos en la toma, organización, recopilación, presentación y análisis de datos, tanto para la deducción de conclusiones como para tomar decisiones razonables de acuerdo con tales análisis.

II. Población y Muestra

A. Población: Es el conjunto de todos los elementos de un grupo de individuos u objetos. La población se llama también universo.

Ejemplo: Población de votantes en Venezuela = conjunto de venezolanos de edad mayor a -x- años.

B. Muestra: Una pequeña parte de la población o subconjunto del universo, seleccionado para estudiar características del grupo.

III. Estadística Descriptiva e Inductiva.

A. Estadística Inductiva o Inferencial

La parte que trata de las condiciones bajo las cuales son válidas inferencias o conclusiones deducidas acerca de la población, a partir del análisis de una muestra.

B. Estadística Descriptiva o Deductiva

La parte que trata solamente de describir y analizar un grupo,sin sacar conclusiones o inferencias de un grupo mayor.

IV. VARIABLE DISCRETAS Y CONTINUAS

A. VARIABLE CONTINUA: Una que puede tomar cualquier valor entre dos valores dados

B. VARIABLE DISCRETA: Si no es contínua.

V. REDONDEO DE DATOS

a) 22.548 ⟹ 22.55
81.8916 ⟹ 81.89

con 5: número par más próximo

b) 102.895 ⟹ 102.90
85.485 ⟹ 85.48

DISTRIBUCIONES DE FREQUENCIA RELATIVA.

I. FRECUENCIA RELATIVA

Para una clase, la frecuencia relativa se obtiene al dividir la frecuencia de la clase por el total de frecuencia de todas las clases, y se expresa como porcentaje.

Por ejemplo se muestran las frecuencias relativas en la Tabla III.

II. DISTRIBUCION DE FRECUENCIA RELATIVA (TABLA III con 3^{era} columna) O DISTRIBUCION PORCENTUAL.

III. HISTOGRAMAS DE FRECUENCIAS RELATIVAS O HISTOGRAMAS PORCENTUALES.

Se obtienen de los histogramas de frecuencia al cambiar la escala verticial de frecuencia a frecuencia relativa, conservando el mismo diagrama.

IV. POLIGONOS DE FRECUENCIAS RELATIVAS O POLIGONOS PORCENTUALES.

Se obtienen de los polígonos de frecuencia al cambiar la escala verticial de frecuencia a frecuencia relativa conservando exactamente el mismo diagrama.

DISTRIBUCIONES DE FRECUENCIA

I. Toma y Ordenación de Datos

II. DISTRIBUCIONES DE FRECUENCIA

1) Distribuir los datos en clases o categorías (intervalos de clase)

2) Determinar frecuencia de cada clase o número de individuos o elementos pertenecientes a cada clase.

3) Distribución de frecuencias o Tabla de Frecuencia:una ordenación tabular de los datos en clases, reunidas las clases, y con las frecuencias correspondientes a cada una

Intervalo de clase	Frecuencia
5-24.9	116
25-49.9	82
50-74.9	49
---------	--------

III. INTERVALOS DE CLASE Y LIMITES DE CLASE

A. Intervalo de clase Ej.: 5-24.9

B. Límites de clase Ej.: 5 y 24.9

1. Límite inferior: 5

2. Límite superior: 24.9

C. Límites reales de clase: 4.95 - 24.95
24.95 - 49.95

D. TAMAÑO o ANCHURA DE UN INTERVALO DE CLASE.

1. Es la diferencia entre los límites reales de clase del intervalo.
2. Se llama también: anchura de clase, longitud de clase o tamaño de clase.
3. Si todos los intervalos tienen igual tamaño, este tamaño o anchura se designa por c.

E. MARCA DE CLASE

1. Es el punto medio del intervalo: para el intervalo 5 - 24.9 la marca es

$$\frac{5 + 24.9}{2} = 14.95$$

2. Se llama también punto medio de la clase

III. REGLAS GENERALES PARA FORMAR DISTRIBUCIONES DE FRECUENCIA

1. Determinar el rango de los datos registrados
2. Dividir el rango en un número conveniente de intervalos de clase del mismo tamaño.
   a) Si esto no es posible utilizar intervalos de clase del $\neq$ tamaño o intervalos de clase abiertos.
   b) Pautas para seleccionar los intervalos de clase:
      i) El número de intervalos se toma entre 5 y 20 dependiendo de los datos.

ii) Elegir los intervalos de forma que las marcas de clase o puntos medios coincidan con datos realmente observados. Esto aminora el llamado error de agrupamiento.

iii) Los límites de clase no deben coincidir con los datos observados.

3. Determinar el número de observaciones que caen dentro de cada intervalo de clase, es decir, encontrar las frecuencias de clase. Lo mejor es utilizar una hoja de conteo.

EJEMPLOS

ILUSTRACION DISTRIBUCION FRECUENCIA

Datos de Escorrentía del Río Paguey en El Paso.

SOLUCION

Para el período de Diciembre a Abril 1950 a 1969 (100 datos)

Escorrentía mínima: 5.0 m^3/s
Escorrentía máxima: 92.3 m^3/s

Rango: 92.3 - 5.0 = 87.3

Si escogimos 20 intervalos de clase, el tamaño de cada intervalo es:

$$\frac{87.3}{20} = 4.36 \simeq 4.5$$

Si, además, escogimos como marcas de clase 7.0, 11.5, 16.0,... entonces

TABLA I. Distribución de frecuencia con tamaño de clase de 4.5

ESCORRENTIA (m^3/s) (Intervalos de Clase)	CONTEO	FRECUENCIA
4.75 - 9.25	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ////	29
9.25 - 13.75	~~////~~ ~~////~~ ~~////~~ ~~////~~ //	22
13.75 - 18.25	~~////~~ ~~////~~ /	11
18.25 - 22.75	~~////~~ ~~////~~ /	11
22.75 - 27.25	~~////~~ /	6
27.25 - 31.75	//	2
31.75 - 36.25	~~////~~	5
36.25 - 40.75	///	3
40.75 - 45.25	//	2
45.25 - 49.75	///	3
49.75 - 54.25	///	3
54.25 - 58.75		0
58.75 - 63.25		0
63.25 - 67.75		0
67.75 - 72.25		0
72.25 - 76.75	//	2
76.75 - 81.25		0
81.25 - 85.75		0
85.75 - 90.25		0
90.25 - 94.75	/	1

TOTAL 100

1. En la Tabla I se presenta una distribución de frecuencia con una anchura de clase c=4.5 Con las ventajas e inconvenientes siguientes

   a) Los límites de clase no coinciden con los datos observados, de modo que no hay confusión en ubicar los datos en los intervalos de clase.

   b) Las marcas de clase coinciden con los datos observados

   c) La información es más detallada en el extremo superior de la escala de escorrentía.

   d) Hay muchas clases vacías.

TABLA II

ESCORRENTIA (m^3/s) (intervalo de clase)	FRECUENCIA
4.75 - 9.25	29
9.25 - 13.75	22
13.75 - 18.25	11
18.25 - 22.75	11
22.75 - 27.25	6
27.25 - 31.75	2
31.75 - 36.25	5
36.25 - 40.75	3
40.75 - 45.25	2
45.25 - 49.75	3
49.75 - 54.25	3
54.25 ó más	3

2. En la Tabla II se presenta una distribución de frecuencia con un intervalo de clase abierto (54.25 ó más), que presenta las ventajas e inconvenientes siguientes:

   a) evita el inconveniente de las clases vacías

   b) presenta el inconveniente de no poder ser útil para determinado cálculos matemáticos. Por ejemplo:

i) no se puede calcular una marca de clase para invervalo abierto (54.25 o más).

ii) no se puede calcular la escorrentía total durante el período, ya que la clase (54.25 o más) no pone límite superior a los valores de la escorrentía en esa clase.

iii) representación gráfica difícil.

3. Otra manera de eliminar las clases vacías será de utilizar un tamaño de clase constante pero mayor. Lo que tendría los efectos siguientes

a) La información es menos precisa para valores bajos de escorrentía, valores en los cuales estamos a menudo interesados.

b) La información es más detallada para valores superiores o grandes de la escorrentía.

c) Error de agrupamiento mayor con mayor tamaño de clase.

4. Otra solución posible consiste en el empleo de intervalos de clase desiguales tal como se ilustra en la Tabla III. En esta tabla los intervalos de clase desiguales son identificados con un asterísco. Características:

a) Evita clases vacías.

b) Se pierde la sencillez que, en ciertos cálculos matemáticos se tiene cuando el intervalo de clase tiene siempre el mismo tamaño.

c) El error de agrupamiento será mayor en el intervalo de mayor tamaño

d) La representación gráfica de la distribución de frecuencia.

TABLA III Distribución de frecuencia con intervalos de clase desiguales*

ESCORRENTIA (m^3/s) Intervalos de Clase)	FRECUENCIA	FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA
4.75 - 9.25	29	29%	29
9.25 - 13.75	22	22%	51
13.75 - 18.25	11	11%	62
18.25 - 22.75	11	11%	73
22.75 - 27.25	6	6%	79
27.25 - 31.75	2	2%	81
31.75 - 36.25	5	5%	86
36.25 - 40.75	3	3%	89
40.75 - 45.25	2	2%	91
45.25 - 49.75	3	3%	94
* 49.75 - 58.75	3	3%	97
* 58.75 - 76.75	2	2%	99
* 76.75 - 94.75	1	1%	100

TOTAL 100

se hace menos sencilla.

DISTRIBUCION DE FRECUENCIA ACUMULADA

I. FRECUENCIA ACUMULADA

La frecuencia total de todos los valores menores que el límite real superior de clase de un intervalo de clase dado, se conoce como frecuencia acumulada hasta este intervalo de clase inclusive. Ilustración: 4 columna Tabla III:

Por ejemplo para el intervalo de clase 27.25 - 31.75 la frecuencia acumulada es 81, significando que 81 valores de escorrentía son inferiores a 31.75 m^3/s

II. DISTRIBUCION DE FRECUENCIA ACUMULADA.

A. TABLA IV. DISTRIBUCION ACUMULADA (menor que)

ESCORRENTIA m^3/s)	FRECUENCIA ACUMULADA menor que	FRECUENCIA RELATIVA
Menor que 4.75	0	
Menor que 9.25	29	29%
Menor que 13.75	51	51%
Menor que 18.25	62	62%
Menor que 22.75	73	73%
Menor que 27.25	79	79%
Menor que 31.75	81	81%
Menor que 36.25	86	86%
Menor que 40.75	89	89%
Menor que 45.25	91	91%
Menor que 49.75	94	94%
Menor que 58.75	97	97%
Menor que 76.75	99	99%
Menor que 94.75	100	100%

B. TABLA V. DISTRIBUCION ACUMULADA (O MAS)

ESCORRENTIA(m^3/s)	FRECUENCIA ACUMULADA O MAS
4.75 o más	100
9.25 " "	71
13.75 " "	49
18.25 " "	38
22.75 " "	27
27.25 " "	21
31.75 " "	19
36.25 " "	14
40.75 " "	11
45.25 " "	9
49.75 " "	6
58.75 " "	3
76.75 " "	1
94.25 " "	0

HISTOGRAMAS DE FREQUENCIA

Son dos representaciones gráficas de las distribuciones de frecuencias

A. HISTOGRAMA o HISTOGRAMA DE FRECUENCIA

Consiste en una serie de rectángulos que tienen:

1. Sus bases sobre el eje horizontal con centro en las marcas de clase y longitud igual al tamaño de los intervalos de clase.

2. Superficies proporcionales a las frecuencias de clase.
   a) Si todos los intervalos tienen igual tamaño, las alturas de los rectángulos son proporcionales a las frecuencias y se acostumbra en tal caso a tomar las alturas numéricamente iguales a las frecuencias de clase.
   b) Si los intervalos de clase no son de igual tamaño, estas alturas deberán ser calculadas como se ilustra a continuación

EJEMPLO

Construir un histograma para la distribución de frecuencias de la Tabla III en donde los intervalos de clase son desiguales.

SOLUCION

1. Para los intervalos de clase de tamaño igual a 4.5 trazamos los rectángulos de alturas iguales a las frecuencias correspondientes.
2. Para el intervalo 49.75-58.75 de tamaño igual a 9 y frecuencia igual a 3, obtendremos la altura del rectángulo correspondiente de la manera siguiente:
   a) Puesto que la superficie bajo el rectángulo correspondiente al intervalo 45.25-49.75 de tamaño 4.5 y frecuencia 3, tiene una superficie de 3(4.5)=13.5,

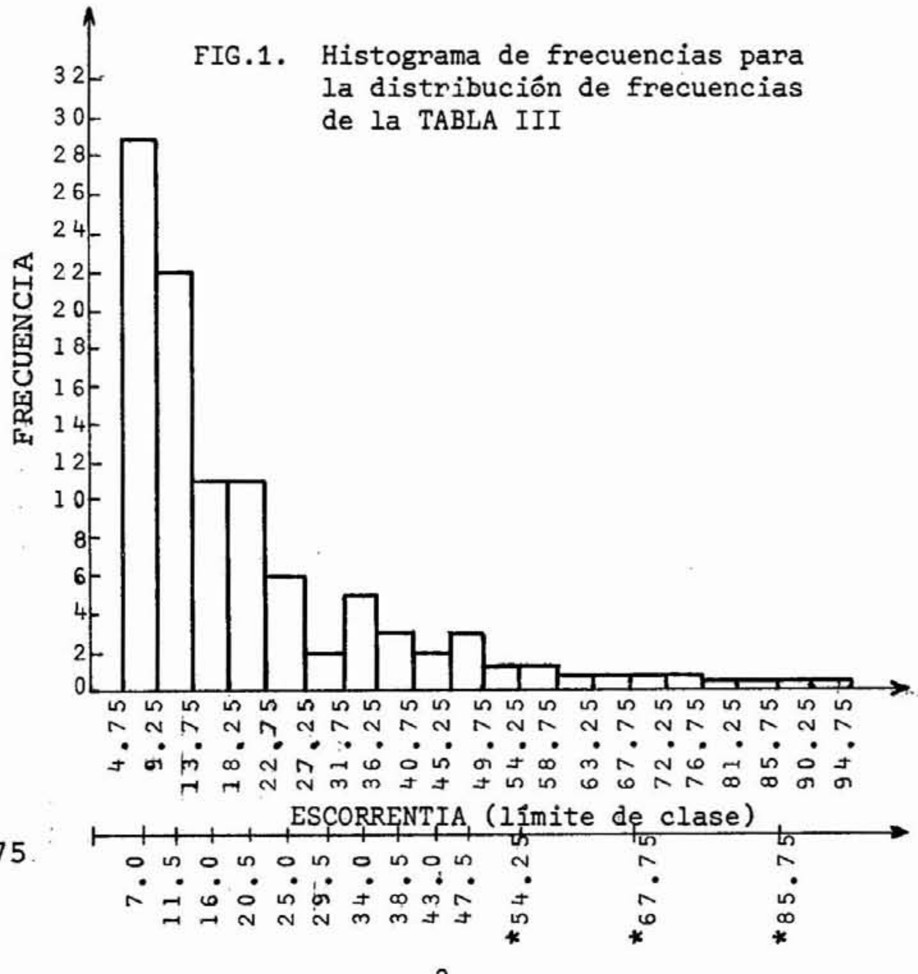


FIG.1. Histograma de frecuencias para la distribución de frecuencias de la TABLA III

el rectángulo correspondiente a la frecuencia 3 y al intervalo 49.75 - 58.75 de tamaño 9 debe tener la misma superficie de 13.5. Como este último rectángulo tiene una anchura 58.75-49.75=9 y una área de 13.5, su altura debe ser:

$$\text{altura} = \frac{\text{área}}{\text{anchura}} = \frac{13.5}{9} = 1.50$$

3. El área del rectángulo correspondiente a un intervalo de tamaño 4.5 y frecuencia 2 (por ejemplo los intervalos 27.25-31.75 y 40.75-45.25) tiene el valor de:

$$4.5 \times 2 = 9$$

El área del rectángulo correspondiente al intervalo 58.75-76.75 de tamaño 18 debe también tener el valor de 9. Como este intervalo tiene una anchura de 76.75-58.75=18, su altura debe ser igual a

$$\frac{9}{18} = 0.50$$

4. Un intervalo de frecuencia 1 y tamaño 4.5 tendría un área de 1x4.5 = 4.5 .

El intervalo 76.75-94.75 debe tener un rectángulo de área igual a 4.5. Como su anchura es igual a 94.75-76.75=18, su altura debe ser igual a:

$$\text{altura} = \frac{\text{área}}{\text{anchura}} = \frac{4.5}{18} = 0.25$$

POLIGONOS DE FRECUENCIA

A. Es un gráfico de línea trazado sobre las marcas de clase. Puede obtenerse uniendo los puntos medios de los techos de los rectángulos en el histograma.

El polígono de frecuencia se ilustra aquí con los datos de la Tabla III o figura 1.

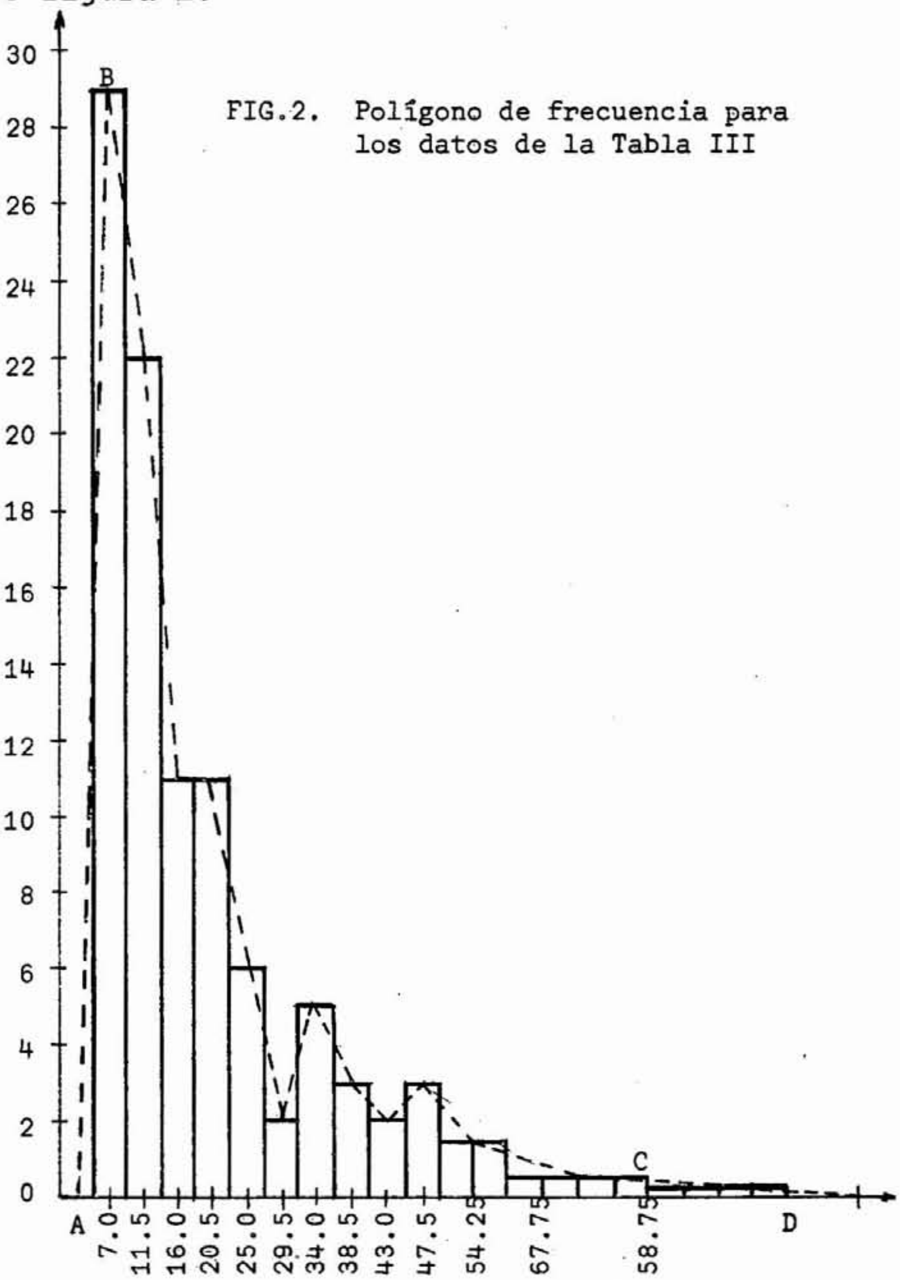


FIG.2. Polígono de frecuencia para los datos de la Tabla III

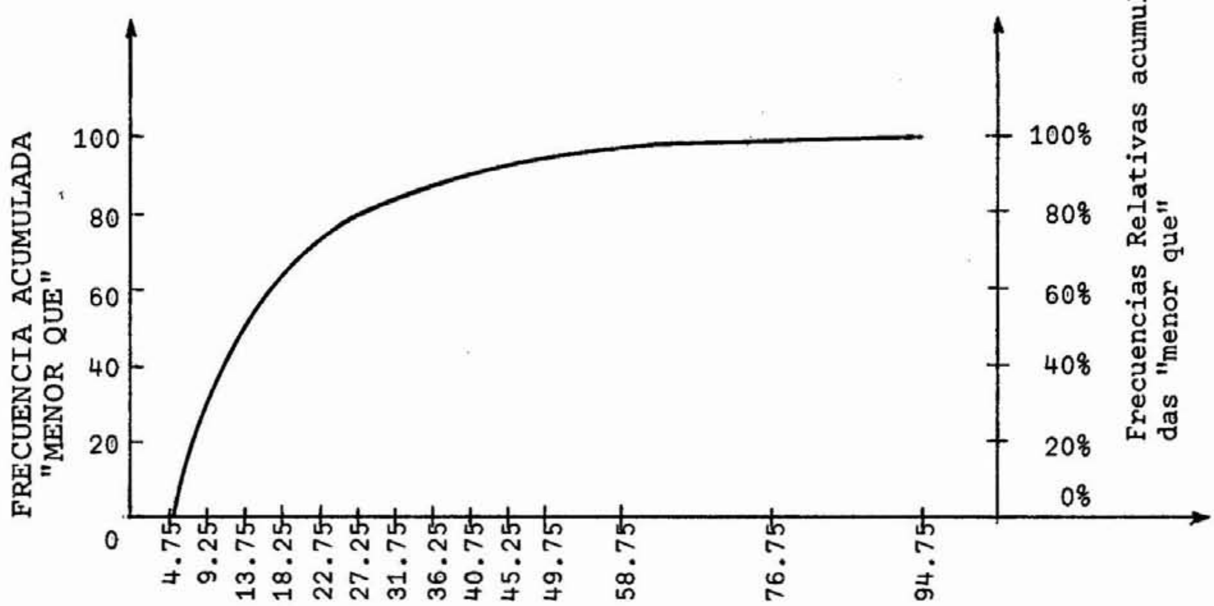
Para obtener el polígono de la figura 2, seguimos los siguientes pasos:

1. Trazamos el histograma de frecuencia tal como lo hicimos en la figura 1.
2. Identificamos los puntos medios de los techos de los rectángulos. Estos puntos corresponden a las marcas de clases.
3. Juntamos cada dos puntos medios sucesivos por una línea recta para obtener el polígono de frecuencia.
4. Se acostumbra a juntar a prolongar el polígono con AB y CD hasta las marcas de clase inferior y superior inmediatas, siendo estas clases de tamaños respectivamente iguales a los intervalos extremos. Por ejemplo el tamaño del intervalo de marca de clase A es igual al del intervalo de marca B, y el tamaño del intervalo de marca de Clase D es igual al del intervalo de marca de Clase C. En tal caso se puede demostrar que el área total de los rectángulos del histograma es igual al área limitada por el polígono de frecuencia.

III. POLIGONOS DE FRECUENCIAS ACUMULADOS U OJIVA

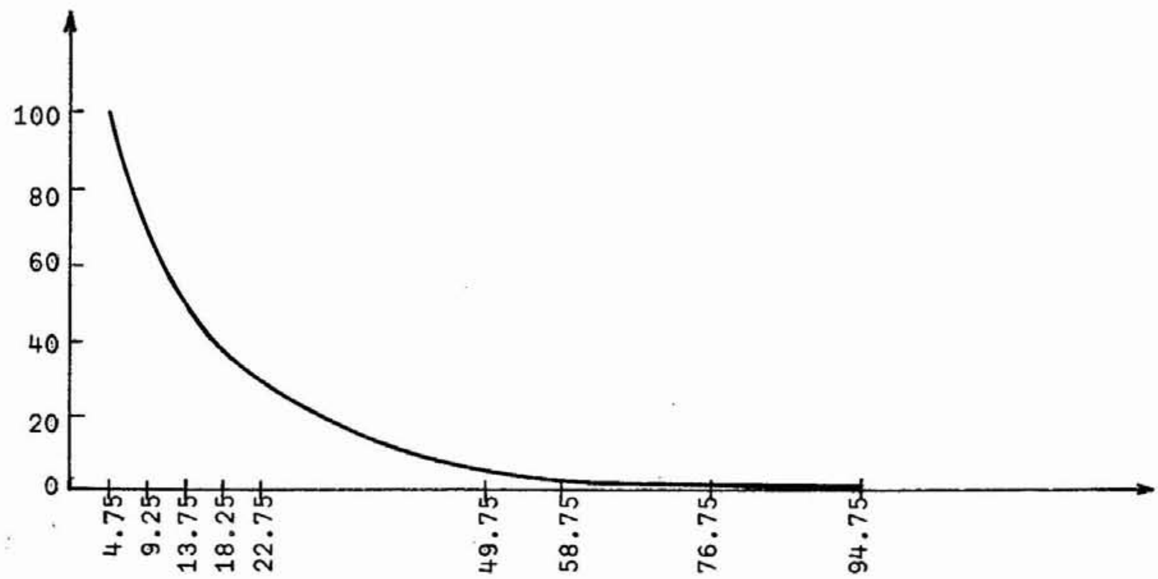
A. Para datos Tabla IV obtenemos.

TABLA VI. POLIGONO DE FRECUENCIA ACUMULADA "MENOR QUE"



B. Para los datos de la Tabla V obtenemos:

TABLA VII. POLIGONO DE FREQUENCIA ACUMULADA O MAS



DISTRIBUCIONES TEORICAS DE PROBABILIDAD EN HIDROLOGIA.

I. DISTRIBUCION NORMAL (o Gausiana)

Funsión de densidad:

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

ó

$$N(u,\sigma)=f(n)$$

Si

$$z= \frac{x-\mu}{\sigma}$$

$$\therefore$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

1. Propiedades

(a) Simétrica

(b) $P(x= \mu-\sigma) = 0.1587$

$P(x = \mu) = 0.50$ de $\overline{x}$

$P(x = \mu+\sigma) = 0.8413$

(c) media $\mu \simeq \overline{x}$

(d) desviación típica: $\sigma \simeq s = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(x_i-\overline{x})^2}$

2. Areas de aplicación en hidrología

(a) Ajuste de distribución empíricas -datos hidrológicos para períodos largos de 2 años, 5 años,......

(b) Análisis de errores aleatorios

(c) Referencia para comparar varias distribuciones.

(d) Para hacer inferencia

(e) Generación de datos (Monte Carlos método)

3. Ajuste:

a) datos para períodos largos > 1 año.

b) dificultades: va de $-\infty$ a $+\infty$

c) ajuste gráfico o analítico, o con factor de frecuencia

d) ajuste mediante papel de probabilidad.

II. DISTRIBUCION LOGNORMAL CON DOS PARAMETROS

A. Función de densidad:

$$f(x) = \frac{1}{x\sigma_n \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu_n}{\sigma_n}\right)^2\right] \quad \text{para } x \geq 0$$

$$f(x) = 0 \qquad \text{para } x < 0$$

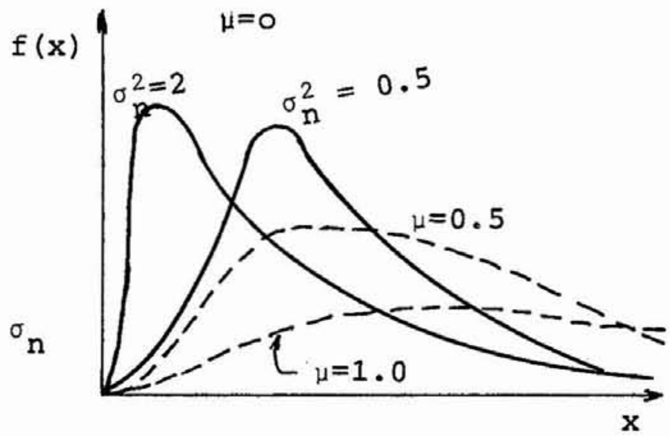
$\mu_n \longrightarrow$ media de $\ln x$ y σ_n desviación típica de $\ln x$

B. Propiedades

1. media : (parámetro) $\mu_n = \frac{1}{2} \ln \left(\frac{\mu^4}{\mu^2+\sigma^2}\right)$ donde μ=media de x

σ^2=varianza de x

2. varianza: (parámetro) $\sigma_n^2 = \ln \left(\frac{\sigma^2 + \mu^2}{\mu^2}\right)$



3. $P(x)= 50\%$ da $\bar{x}_n$

$P(x_1)= 84.13\ \%$

ó da $s_n \simeq \sigma_n$

$P(x_2)= 15.87\ \%$

4. Sesgada C_s=1.139 y C_v=0.364

C. Aplicaciones en Hidrología

1. Se ajusta mejor a distribuciones empíricas de escorentía y precipitación anuales, y de dimensiones de gramos de sedimentos.

2. Valores de 0 a + ∞ de variables hidrológicas.

D. AJUSTE

1. Por método de verosimilitud máxima requiere cálculo de los lnx
2. Métodos de momentos y gráficos más sencillos pero menos precisos.

3. Existe papel de probabilidad
4. Mediante factor de frecuencia.

III. DISTRIBUCION LOGNORMAL CON TRES PARAMETROS

Si el límite inferior de la variable hidrológica no es cero, se hace necesario modificar **la distribución** lognormal al introducir el límite inferior x_o como tercer parámetro.

A. FUNCION DE DENSIDAD

$$f(x) = \frac{1}{(x-x_o)\,\sigma_n\,\sqrt{2\pi}}\,\exp\left\{-\frac{\left[\ln(x-x_o)-\mu_n\right]^2}{2\,\sigma_n^2}\right\}$$

para $x_o \leq x \leq +\infty$

$f(x) = o$ para $x < x_o$

B. PROPIEDADES

1. Media (1°parámetro) μ_n de $\ln(x-x_o)$. Se puede calcular de:

$$\mu = x_o + \exp\left(\frac{\sigma_n^2}{2} + \mu_n\right) \qquad 1)$$ donde $\mu \rightarrow$ media de x

2. Varianza σ_n^2 de ln $(x-x_o)$ 2°parámetro se puede calcular de:

$$\sigma^2 = \left[\exp(\sigma_n^2) - 1 \right] \exp(2\,\mu_n + \sigma_n^2) \qquad 2)$$

donde σ —> desviación típica de x.

3. Límite inferior x_o (3°parámetro).

Si se conoce x_o de antemano se puede calcular μ y σ^2 de (1) y (2). Sin embargo, esto no es generalmente el caso y se tiene que estimar también a x_o

a. Un método presentado por Biswas, es de utilizar la mediante γ de x y calcular x_o mediante:

$$x_o = \gamma - \frac{\sigma^2}{2(\mu - \gamma)}$$

b. Otro método es de hacer uso del estadístico γ coeficiente de sesgo de las n :

$$\gamma = \frac{\exp(3\,\sigma_n^2) - 3\exp(\sigma_n^2) + 2}{\left[\exp(\sigma_n^2) - 1\right]^{3/2}}$$

C. LEY DE GIBRAT

Otra forma de la distribución log-normal con tres parámetros es la propuesta por R. Gibrat.

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{u^2}{2}} du$$

donde

$$u = a \log (x - x_o) + b$$

D. Otra forma:

$$F(x) = \frac{1}{\sigma_n \sqrt{2\pi}} \int_{o}^{u} \frac{1}{u} \exp \left(\frac{-\log^2 u}{2\sigma_n}\right) du$$

donde

$$u = \frac{x - x_o}{s}$$

1. Propiedades

a. x_o —> parámetro de posición: límite inferior

b. s —> parámetro de escala, positivo y diferente de cero

c. σ —> parámetro de forma: desviación típica de u

2. Estimación de parámetros (ver Ch O.R.S.T.O.M. sér. Hydrol. vol. VI, N°3, 1969)

3. Aplicación en Hidrología

- Una serie anual de caudal tiene su valor mínimo superior a cero. Consecuentemente, debe existir un caudal > 0 de bajo del cual la probabilidad es cero o sea:

$$F(x) = 0 \quad \text{para } x = x_o$$

$$F(x) = 1 \quad \text{para } (x-x_o) \longrightarrow +\infty$$

de modo que se aplica mejor a valores máximos de caudales (Csoma, publication N°84 AIHS)

IV DISTRIBUCION GAMMA CON DOS PARAMETROS

A. Función de densidad

$$f(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\frac{x}{\beta}} \qquad \text{para } o \leq x \leq +\infty$$

$$f(x) = o \qquad \text{para} \quad x < o$$

B. Propiedades

1. $\alpha \longrightarrow$ parámetro de forma $\alpha > o$

$\beta \longrightarrow$ parámetro de escala $\beta > o$

El valor esperado y la varianza es

$E(x) = \mu = \alpha\ \beta$

$\sigma^2(x) = \alpha\ \beta^2$

sesgo: $\gamma_s = \frac{2}{\sqrt{\alpha}}$

curtosis: $\gamma_c = 3 + \frac{6}{\alpha}$

la moda: $m = \beta^{\alpha}\ (\alpha - 1)$ para $\alpha \geq 1$

C. Aplicación en Hidrología

- Similar a la distribución lognormal con 2 parámetros y no $\neq$ cia
- Precipitación y caudal anual
- pero no ajuste gráfico (recta)

N.B. $\Gamma(\alpha) = (\alpha - 1)$ si α es un entero positivo

si $\alpha > o$ y no es un entero: $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}\ e^{-x}\ dx$

V. DISTRIBUCION GAMMA CON TRES PARAMETROS

A. Función de Densidad:

$$f(x) = \frac{1}{\beta^{\alpha} \; \Gamma(\alpha)} (x-\gamma)^{\alpha-1} e^{-(x-\gamma)/\beta}$$

$$\gamma \leq x > +\infty$$

$$\gamma \neq 0$$

B. Propiedades

1. $\alpha \longrightarrow$ parámetro de forma

2. $\beta \longrightarrow$ parámetro de escala

3. $\gamma \longrightarrow$ parámetro de posición

$$E(x) = \mu = \gamma + \alpha\,\beta$$

$$\sigma^2(x) = \alpha\,\beta^2$$

sesgo: $\gamma_s = 2/\sqrt{\alpha}$

curtosis: $\gamma_c = 3 + 6/\alpha$

moda: $m = \gamma + \beta^{\alpha}(\alpha - 1)$

C. Aplicación en Hidrología.

- Similar a la distribución lognormal con 3 parámetros (Csoma)
- Uso para caudales máximos.

UNIVERSIDAD DE LOS ANDES

CIDIAT

CENTRO DE COMPUTACION HIBRIDA

PROYECTO ALTO-APURE (CIDIAT-CORPOANDES)

	ESTACION PATRON		
NOMBRE DE LA ESTACION	******PE******	**EDO**	****INST.****
PAGUEY EL PASO	******OP******	**BAR**	****02-44****

AÑO	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC	PROM.ANUAL
50	44.0	8.3	12.2	15.6	57.1	42.8	61.4	76.4	53.2	70.8	38.6	27.2	42.3
51	13.5	49.7	15.6	16.1	24.9	60.9	69.7	58.2	48.0	29.0	21.3	11.3	34.8
52	5.8	5.8	6.3	46.3	55.5	89.4	119.0	80.0	68.9	61.4	37.2	20.3	49.7
53	10.9	9.7	9.2	36.2	73.6	81.5	74.5	83.3	85.0	64.5	49.7	16.0	49.5
54	8.8	7.8	6.7	46.6	76.0	85.8	92.2	84.5	88.0	75.4	52.8	30.9	54.7
55	12.6	7.8	9.0	33.4	37.6	74.2	122.0	60.9	90.0	66.5	62.9	21.9	49.9
56	22.7	14.8	23.8	33.9	79.0	56.4	55.5	95.1	80.0	50.7	41.0	25.8	48.2
57	10.9	9.2	7.2	19.2	55.5	74.5	112.0	72.9	61.6	43.4	42.4	24.5	44.4
58	11.9	6.8	8.5	17.8	58.4	126.0	80.9	94.1	60.2	48.9	39.6	14.8	47.3
59	7.9	5.0	17.1	40.1	105.0	120.0	83.4	75.5	118.0	80.6	47.3	22.4	60.1
60	13.2	92.3	76.0	42.0	71.7	110.0	114.0	124.0	101.0	35.2	34.2	26.0	69.9
61	9.5	7.1	11.3	8.9	27.9	82.6	75.8	74.4	112.0	114.0	91.7	34.3	54.1
62	14.8	9.8	10.7	38.1	47.8	109.0	93.0	122.0	87.9	85.6	35.0	15.8	55.7
63	11.9	6.8	6.6	22.4	89.5	98.8	95.8	79.7	100.0	56.5	58.0	18.4	53.5
64	9.4	6.6	6.1	18.9	46.8	71.4	104.0	115.0	82.0	55.3	32.7	29.9	48.1
65	18.5	11.8	74.8	13.7	47.2	89.6	52.4	73.0	99.0	85.6	56.9	19.7	53.5
66	12.7	8.7	7.5	21.5	46.0	80.1	71.5	83.9	76.9	60.0	83.1	52.1	50.3
67	14.3	9.3	6.7	50.2	54.9	97.8	98.0	104.0	89.0	79.9	38.9	25.3	55.6
68	11.6	7.4	7.8	40.2	50.2	12.0	87.5	73.3	88.8	63.4	29.2	11.9	40.2
69	8.0	8.7	12.7	51.5	55.5	77.1	72.4	90.0	78.8	114.0	73.0	33.5	56.2
PROMEDIOS	13.6	14.6	16.7	30.6	58.0	81.9	86.7	86.0	83.5	67.0	48.1	24.1	50.9

\- DATOS FALTANTES (CALCULADOS POR INTERPOLACION)

\+ DATOS CALCULADOS POR CORRELACION (MINIMOS CUADRADOS)

TABLA 9. Datos de escorrentía medio mensual del río Paguey en El Paso.