

# ANÁLISIS DE FRECUENCIAS

## EXPRESIONES PARA EL CÁLCULO DE LOS EVENTOS PARA EL PERÍODO DE RETORNO $T$ Y DE LOS RESPECTIVOS ERRORES ESTÁNDAR DE ESTIMACIÓN REQUERIDOS PARA LA DETERMINACIÓN DE LOS INTERVALOS DE CONFIANZA DE LOS ESTIMADOS DE LOS VALORES ESPERADOS

JULIAN DAVID ROJO HERNÁNDEZ

### 3.1 EXPRESIÓN GENERAL PARA EL CÁLCULO DEL ESTIMADO DEL VALOR ESPERADO DE UN EVENTO PARA UN PERÍODO DE RETORNO $T$ (expresión de Ven T. Chow para análisis de frecuencias de eventos extremos):

Sea la serie de tiempo  $X$ :

$$X = \{x_1, x_2, x_3, \dots, x_N\}$$

la muestra observada de los  $N$  valores medidos de un fenómeno o evento cualquiera (lluvias, caudales, temperaturas, sismos, etc.), tiene las siguientes características estadísticas (muestrales):

$\bar{X}$ : estimado (muestral) del valor esperado del fenómeno analizado  $X$ , equivalente al valor medio o promedio de los  $N$  valores observados de la serie  $X$ ; se le llama también "media" de  $X$ . Corresponde al denominado primer momento muestral del arreglo  $X$  con relación al origen de los reales (con respecto a cero).

$S_x$ : estimado (muestral) de la desviación típica (desviación estándar) muestral (raíz cuadrada de la varianza), obtenido a partir de la utilización de los  $N$  valores observados de la serie  $X$ , empleando el estimador insesgado de este parámetro. Corresponde a la raíz cuadrada del segundo momento muestral del arreglo  $X$  con respecto de la media.

$g_x$ : estimado del coeficiente de asimetría muestral (raíz cuadrada de la varianza), obtenido a partir de la utilización de los  $N$  valores observados de la serie  $X$ , empleando el estimador insesgado de este parámetro. Corresponde a la relación entre el tercer momento muestral del arreglo  $X$  con relación a la media, y el cubo del estimado sesgado de  $S_x$ .

Debido a la incertidumbre en los estimados de parámetros estadísticos muestrales de orden superior al tercer momento muestral, no se tienen en cuenta parámetros como el de kurtosis y similares.

Si los valores de la serie X son eventos aleatorios independientes entre sí, pertenecientes a una misma función de distribución (densidad) de probabilidades  $f(x; \alpha, \beta, \gamma, \dots)$ , tal que  $\alpha, \beta, \gamma, \dots$  son los parámetros de la función de distribución de probabilidades, se puede obtener los valores estimados de la variable X para cualquier probabilidad de ocurrencia dada  $p$  a partir de la siguiente expresión general (propuesta por Ven T, Chow):

$$X_p = \bar{X} + K_p \cdot S_X$$

Donde  $K_p$  es un número adimensional, aleatorio, obtenido de la función de distribución de probabilidades cualquiera  $f(x; \alpha, \beta, \gamma, \dots)$  a la que pertenecen las ocurrencias aleatorias del fenómeno X.

Como es sabido, en el manejo ingenieril tradicional de las probabilidades de ocurrencia de fenómenos, el concepto de definición de la probabilidad de acurrencia ha sido asociado al concepto de período de retorno o recurrencia del evento en consideración (T), el cual, desde el punto de vista de su formalización matemática ha sido definido de la siguiente forma:

$$T = 1/p$$

Así, cuando se habla de eventos extremos máximos, la probabilidad de interés es la probabilidad de excedencia de un evento ( $G(x) = 1 - F(x)$ ), mientras cuando se trata de eventos extrmos mínimos dicha probabilidad de interés es la de no excedencia ( $F(x) = P[X \leq x]$ ). Por esta razón, aunque para algunos pueda ser más ilustrativo utilizar el concepto del “período de retorno”, su empleo en el manejo numérico de probabilidades debe ser consecuente para evitar errores de cálculo graves.

De esta forma, en la práctica, cuando se habla de períodos de retorno para el caso de eventos máximos, la expresión de T es el inverso de la probabilidad de excedencia del evento de interés, o sea:

$$T = 1/G(x) = 1/[1 - F(x)]$$

mientras que si se trata de períodos de retorno para el caso de eventos mínimos, la expresión de T es el inverso de la probabilidad de no - excedencia del evento de interés, o sea:

$$T = 1/F(x)$$

Con estas advertencias, la forma más popular de la expresión de Ven T. Chow para el análisis de frecuencias (de máximos o de mínimos) es:

$$X_T = \bar{X} + K_T \cdot S_X \quad (1)$$

El valor de  $X_T$  obtenido con la expresión (1) corresponde al estimado del valor esperado del evento X para el período de retorno T, el cual se encuentra exactamente dentro de la función matemática que (se supone) describe el comportamiento probabilístico de la variable X que se está considerando.

### 3.2 INTERVALO DE CONFIANZA PARA EL ESTIMADO DE $X_T$

Como se ha dicho en clase,  $X_T$  de la expresión (1) es solo el valor esperado del evento X para el período de retorno T, ya debe recordarse que realmente  $X_T$  es una variable aleatoria a causa, entre otras, de la incertidumbre originada en la estimación de los parámetros de la distribución de probabilidades de X (y en su selección, por supuesto), de manera que pueden existir (son probables) valores de X diferentes al calculado con la expresión (1), el cual se supone que es simplemente el valor esperado de X para ese T; es más, ni siquiera puede afirmarse que el resultado de la expresión (1) sea el valor más probable, ya que la distribución de probabilidades de los valores de X para dicho T (distribución marginal de X para T) no necesariamente es simétrica (en cuyo caso el valor esperado sí es a la vez el más probable). Por lo tanto, el valor  $X_T$  tiene incertidumbre, la cual se calcula a partir de la estimación de la varianza de X para dicho período de retorno T, cuya raíz cuadrada es conocida como “Error Estándar”,  $S_E(T)$ .

En términos generales,  $S_E(T)$  es función de  $S_X$ , T, N y la  $f(x; \alpha, \beta, \gamma, \dots)$  a la que se supone pertenece X.

De esta forma, conocidos  $X_T$  y el correspondiente valor de  $S_E(T)$  (para el mismo T), es posible calcular el intervalo de confianza del estimado del evento X para el período de retorno T y un nivel de significancia  $\alpha$ , de acuerdo con la siguiente expresión:

$$(X_T + \omega_{\alpha/2} \cdot S_E(T)) < X_T < (X_T + \omega_{1-\alpha/2} \cdot S_E(T)) \quad (2)$$

donde  $\omega_{\alpha/2}$  ó  $\omega_{1-\alpha/2}$  son respectivamente los valores (adimensionales) representativos de las probabilidades de no excedencia  $\alpha/2$  y de excedencia  $1 - \alpha/2$  de la función de distribución de probabilidades marginal de los eventos de la variable X para el período de retorno T, la cual tiene como valor medio  $X_T$  y como desviación típica  $S_E(T)$ .; naturalmente,  $\omega_{\alpha/2}$  y  $\omega_{1-\alpha/2}$  pueden ser positivos o negativos, según el nivel de probabilidad  $\alpha/2$  y  $1 - \alpha/2$  que sea definido.

Lo anterior quiere decir que existe una probabilidad  $\alpha$  de que  $X_T$  se encuentre por fuera del intervalo de confianza definido en la expresión (2).

En este caso ( $X_T + \omega_{\alpha/2} \cdot S_E(T)$ ) corresponde al denominado "límite inferior de confianza", y ( $X_T + \omega_{1-\alpha/2} \cdot S_E(T)$ ) al "límite superior de confianza", con una "confiabilidad" de  $(1 - \alpha)$  (nivel de significancia  $\alpha$ ).

Estrictamente hablando, para calcular los valores  $\omega$  se debe conocer la distribución de probabilidades de los valores de  $X$  para el período de retorno  $T$  en consideración (distribución marginal), la cual no obstante es desconocida, por lo cual se utiliza habitualmente la distribución "t" (Student), la cual tiene la ventaja de considerar el número de datos de la muestra  $X$ .

No obstante, algunos investigadores conceptúan que para propósitos prácticos aplicados en hidrología, para el cálculo de los límites de confianza del un evento  $X$  en el período de retorno  $T$  se puede utilizar la distribución de probabilidades normal<sup>(1)</sup>, simplificando de esta manera la utilización de las diferentes herramientas de cálculo, aprovechando la existencia de similares condiciones de incertidumbre y precisión de los resultados.

Teniendo en cuenta esta consideración sobre la distribución probabilística de los valores de  $X$  para el período de retorno  $T$ , los intervalos de confianza para los estimados de un evento  $X$  en el período de retorno  $T$  utilizando la distribución de probabilidades normal, queda de la siguiente forma:

$$\text{Límite inferior del intervalo} = X_T + z_{\alpha/2} \cdot S_E(T)$$

$$\text{Límite superior del intervalo} = X_T + z_{1-\alpha/2} \cdot S_E(T)$$

Donde  $z_{\alpha/2}$  y  $z_{1-\alpha/2}$  corresponden a los valores de la variable Normal estandarizada o tipificada para probabilidades de no excedencia y de excedencia de  $\alpha/2$  y  $1 - \alpha/2$ , respectivamente.

Si se adopta un nivel de significancia del 5% ( $\alpha=0,05$ ) normalmente utilizado en estudios hidrológicos, los límites de confianza quedarán de la siguiente forma:

$$\text{Límite inferior del intervalo} = X_T - 1,96 \cdot S_E(T)$$

$$\text{Límite superior del intervalo} = X_T + 1,96 \cdot S_E(T)$$

### **3.3 EXPRESIONES PARA EL CÁLCULO DE $K_T$ Y $S_E(T)$ PARA DIFERENTES DISTRIBUCIONES DE PROBABILIDADES UTILIZADAS EN HIDROLOGÍA OBTENIDAS CON BASE EN EL MÉTODO DE LOS MOMENTOS**

Se utilizan las expresiones derivadas en diferentes textos de estadística aplicada a la hidrología en el análisis de eventos extremos, obtenidas a partir de la aplicación de técnicas de estimación por el método de los momentos.

---

<sup>(1)</sup> KITE, G. W. Confidence Limits for Design Events. Water Resources Research, Vol 11, No. 1, pp. 48-53. 1975.

Distribución de probabilidades	Expresiones para $K_T$ y $S_E(T)$
Normal	$K_T = z_T$ (*)
	$S_E(T) = S_X \cdot \{ [1 + (z_T)^2/2] / N \}^{1/2}$
Lognormal de dos parámetros (espacio real)	$K_T = \frac{\exp\{z_T \cdot [\ln(1 + C_{vX}^2)]^{1/2} - 0,5 \cdot \ln(1 + C_{vX}^2)\} - 1}{C_{vX}}$
	$S_E(T) = S_X \cdot \{ [1 + (C_{vX}^3 + 3 C_{vX})K_T + (1/4)(C_{vX}^8 + 6 C_{vX}^6 + 15 C_{vX}^4 + 16 C_{vX}^2 + 2)K_T^2] / N \}^{1/2}$
Eventos Extremos	$K_T = -0,45 - 0,7797 \ln \{ -\ln[F(x)] \}$ (**)
Tipo I, o Gumbel	$S_E(T) = S_X \cdot \{ [1 + 1,1396 K_T + 1,1 K_T^2] / N \}^{1/2}$
Pearson Tipo III	$K_T = z_T + (z_T^2 - 1)(g_X/6) + (1/3)(z_T^3 - 6z_T)(g_X/6)^2 - (z_T^2 - 1)(g_X/6)^3 + (z_T)(g_X/6)^4 - (1/3)(g_X/6)^5$
	$S_E(T) = S_X \cdot \{ \{ 1 + g_X K_T + (K_T^2/2)(3 g_X^2/4 + 1) + (3 K_T)(W)(g_X + g_X^3/4) + 3(W^2)(2 + 3 g_X^2 + 5 g_X^4/8) \} / N \}^{1/2}$ (***)

(\*)  $z_T$  : Variable normal tipificada asociada a una probabilidad  $p = 1/T$

(\*\*) Recuérdese que para el caso de análisis de eventos máximos  $p = 1 - F(x) = 1/T$ , y para el análisis de eventos mínimos  $p = F(x) = 1/T$ . Por lo tanto, para el caso de análisis de **máximos**,  $\ln [F(x)] = \ln(1-1/T)$ , mientras que para los eventos **mínimos**  $\ln [F(x)] = \ln(1/T)$

(\*\*\*)  $W = (z_T^2 - 1)/6 + 4 g_X (z_T^3 - 6z_T)/6^3 - 3(g_X^2)(z_T^2 - 1)/6^3 + 4(g_X^3)(z_T)/6^4 - 10(g_X^4)/6^6$

En el caso de las distribución de probabilidades lognormal de dos parámetros, en la tabla anterior se presentan las expresiones correspondientes al cálculo de  $X_T$  y  $S_E(T)$  directamente en el espacio de los valores medidos de la variable X (no transformados). No obstante debe recordarse que para esta distribución el análisis se puede realizar también, en su totalidad, en el espacio de los logaritmos, caso en el cual ya no se utilizan el valor medio y la desviación típica de los valores de X, sino los parámetros estadísticos de sus logaritmos naturales (neperianos); en este caso se utilizan directamente las expresiones mostradas para el caso de la distribución normal, solo que  $X_T$  y  $S_E(T)$  corresponden al valor esperado del estimado del logaritmo natural del evento para el período de retorno T y su correspondiente error estándar (en el campo de los logaritmos, por supuesto), de manera que los intervalos de confianza se determinan en el espacio de los logaritmos, de forma tal que el resultado en el espacio de medición de la variable original X se obtiene a partir del cálculo del antilogaritmo de los límites de confianza antes obtenidos.

En el caso de la distribución de probabilidades lognormal de dos parámetros, la variable  $C_{vx}$  representa el coeficiente de variación de la variable X, el cual se calcula como la relación entre la desviación típica y la media de X.

El caso de la popular distribución de probabilidades Log Pearson Tipo III es similar al de la distribución Lognormal, ya que también bastaría con utilizar las expresiones antes mostradas correspondientes a la distribución Pearson Tipo III, pero aplicada a los logaritmos de la variable de interés, de manera que los antilogaritmos de los resultados así obtenidos corresponden a los resultados de la distribución Log Pearson Tipo III.

**Ejemplo 3.1:** Se desea canalizar la quebrada La Concepción, y para ello se diseñan obras cuya vida útil se estima en 50 años, además se admite un riesgo de daño de la estructura del 50%. Determinar los caudales de diseño para la canalización utilizando las distribuciones de probabilidad Normal, Lognormal, Gumbel y Pearson tipo III, con sus respectivos intervalos de confianza. Para ello considere los datos de Caudales máximos mensuales de la estación Los Siprés ubicada en inmediaciones del sitio de la obra. (Figura 1).

VALORES MAXIMOS MENSUALES DE CAUDALES (m3/seg)																	SISTEMA DE INFORMACION NACIONAL AMBIENTAL											
FECHA DE PROCESO : 2012/04/16										ESTACION : 23087180 SIRPES LOS																		
LATITUD 0622 N										FECHA-INSTALACION 1973-MAY																		
LONGITUD 7509 W										FECHA-SUSPENSION																		
ELEVACION 1571 m.s.n.m																												
TIPO EST 01																												
ENTIDAD 01 IDEAM																												
REGIONAL 01 ANTIOQUIA																												
CORRIENTE QDA LA CONCEPCION																												
AÑO	EST	ENT	ENERO	FEBRE	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOST	SEPTI	OCTUB	NOVIE	DICIE	VR ANUAL													
1973	2	01	19.10	6	5.400	6	38.10	6	23.76	6	31.10	6	25.80	1	58.20	1	67.70	1	45.80	1	34.70	1	70.58	8	23.00	70.58		
1974	2	01	31.75	12.00	57.60	1	38.00	1	48.02	1	67.70	6	90.66	6	77.52	6	59.50	1	75.10	6	79.47	6	34.00	1	90.66	90.66		
1975	2	01	26.15	23.35	1	28.38	1	35.60	1	30.63	49.36	1	61.07	1	82.90	8	80.90	8	61.60	1	44.50	1	40.45	1	82.90	82.90		
1976	2	01	22.35	1	38.35	1	48.91	1	64.22	1	51.15	1	48.91	1	37.20	1	38.35	1	19.75	1	106.0	8	37.60	1	14.15	106.00		
1977	2	01	5.480	1	12.50	1	25.22	1	21.80	1	25.96	1	49.20	1	45.35	1	51.34	1	41.13	1	28.18	1	28.18	1	8.000	1	51.34	
1978	2	01	20.60	1	18.50	1	29.29	1	77.52	1	44.92	1	44.92	1	62.89	1	23.00	1	44.07	1	56.47	1	45.35	1	22.10	77.52		
1979	2	01	5.000	1	16.40	1	23.00	1	11.00	1	30.00	1	33.50	1	25.10	1	24.40	1	12.50	1	47.06	1	32.80	1	12.20	47.06		
1980	2	01	9.200	13.40	8.000	14.00	20.00	33.15	1	44.94	1	13.40	23.00	67.70	8	37.50	54.30	8	67.70	8	37.50	54.30	8	67.70	67.70			
1981	2	01	5.600	9.200	17.00	27.20	58.60	8	30.00	45.80	30.00	24.75	30.70	23.00	18.20	58.60	62.89	20.00	4.600	62.89	20.00	4.600	62.89	62.89	62.89			
1982	2	01	9.500	23.00	17.00	33.50	41.50	9.800	15.20	17.00	20.00	62.89	8	20.00	4.600	62.89	20.00	4.600	62.89	20.00	4.600	62.89	20.00	4.600	62.89			
1983	2	01	5.000	5.800	4.600	33.50	23.00	11.00	23.00	14.00	33.50	27.90	23.70	1	25.10	1	33.50	33.50	27.90	23.70	1	25.10	1	33.50	33.50			
1984	2	01	25.45	1	33.50	1	6.460	1	28.60	1	47.48	1	76.42	8	23.00	43.22	1	50.00	1	26.50	85.29	8	17.00	85.29	85.29			
1985	2	01	17.00	24.40	1	29.30	1	6.240	53.01	8	16.40	1	47.90	1	44.94	1	92.26	8	41.50	1	41.50	1	25.10	1	92.26			
1986	2	01	39.10	1	22.40	1	22.70	1	38.70	1	39.90	1	83.30	8	59.89	8	54.30	8	23.00	1	46.64	1	33.50	1	39.10	1	83.30	
1987	2	01	18.50	1	29.30	1	16.70	1	38.70	1	102.2	8	52.60	8	85.80	8	50.00	1	60.70	8	56.00	8	32.80	1	32.10	1	102.20	
1988	2	01	17.90	1	21.80	1	18.20	3	85.79	8	44.51	1	43.65	3	24.40	1	62.89	8	61.17	8	69.62	8	81.81	3	19.40	1	85.79	3
1989	2	01	57.74	8	23.00	1	35.10	1	44.94	1	53.01	8	61.60	8	47.90	1	48.74	1	65.78	8	53.01	8	39.50	1	95.98	8	95.98	
1990	2	01	19.52	1	41.89	8	55.55	8	41.89	8	59.70	8	54.31	8	52.23	8	56.38	8	44.79	8	40.65	8	60.53	8	48.09	8	60.53	
MEDIOS			19.72	20.79	26.73	36.94	44.71	43.98	47.25	44.45	44.59	51.79	45.42	29.60	38.00													
MAXIMOS			57.74	41.89	57.60	85.79	102.2	83.30	90.66	82.90	92.26	106.0	85.29	95.98	106.00													
MINIMOS			5.000	5.400	4.600	6.240	20.00	9.800	15.20	13.40	12.50	26.50	20.00	4.600	4.600													

Figura 1: Caudales máximos mensuales de la estación Liminigráfica Siprés-Los sobre la Quebrada la Concepción.

**Solución:**

Para iniciar los cálculos se deben seleccionar los caudales máximos anuales ( el valor máximo de los caudales de cada año), y estimar los momentos estadísticos más significativos de la muestra de caudales máximos, ellos son la media, la varianza ( cuya raíz corresponde a la desviación estándar), y el coeficiente de asimetría, además de determinar el coeficiente de variación.

AÑO	$X$	$(X - \bar{X})^2$	$(X - \bar{X})^3$
1973	70.6	21.6	-100.4
1974	90.7	238.2	3675.2
1975	82.9	58.9	451.6
1976	106.0	946.9	29139.1
1977	51.3	570.6	-13631.0
1978	77.5	5.3	12.0
1979	47.1	793.4	-22349.0
1980	67.7	56.7	-426.6
1981	58.6	276.5	-4597.3
1982	62.9	152.2	-1878.1
1983	33.5	1741.2	-72656.7
1984	85.3	101.2	1018.8
1985	92.3	290.1	4941.0
1986	83.3	65.2	526.0
1987	102.2	727.5	19622.3
1988	85.8	111.6	1178.3
1989	96.0	430.7	8937.0
1990	60.5	216.0	-3175.1
$\Sigma$	1354.1	6803.7	-49312.6

Sea  $X$  la muestra observada de los  $N = 18$  caudales máximos de la estación Los Sipres sobre la quebrada Concepción.

Paso 1: siguiendo las ecuaciones de los estimadores insesgados para los principales momentos estadísticos de una muestra aleatoria de datos se tiene que:

Media: 
$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{18}(1354.1) = 75.23 m^3 / s$$

Varianza: 
$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 = \frac{1}{18-1}(6803.7) = 400.22 m^6 / s^2$$

Desviación estándar: 
$$S_x = \sqrt{S_x^2} = \sqrt{400.22} = 20 m^3 / s$$

Coefficiente de asimetría: 
$$g_x = \frac{N \sum_{i=1}^N (X_i - \bar{X})^3}{(N-1)(N-2)(S_x^3)} = \frac{18(-49312.6)}{(18-1)(18-2)(20^3)} = -0.41$$

Coefficiente de variación: 
$$C_{vx} = \frac{S_x}{\bar{X}} = \frac{20}{75.23} = 0.266$$

Paso 2: Para iniciar los cálculos se debe determinar el periodo de retorno necesario para la estimación de las probabilidades de ocurrencia del evento hidrológico de diseño. Para ello se recurre la expresión que relaciona el riesgo  $R$ , con la vida útil  $n$  y el periodo de retorno  $T$ :

$$R = 1 - \left(1 - \frac{1}{T}\right)^n$$

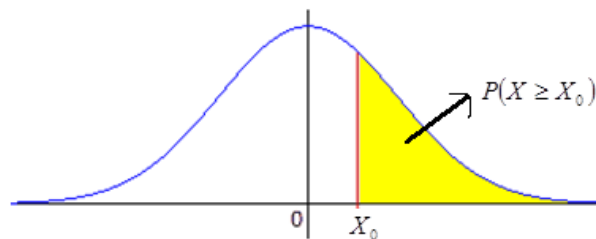
$$0.25 = 1 - \left(1 - \frac{1}{T}\right)^{50}$$

De donde se obtiene que  $T = 72.6 \text{ años}$

Paso 3, por ser un diseño para caudales máximos, la probabilidad de interés corresponde a la probabilidad de excedencia de un evento de referencia  $X_0$  que corresponde al valor del caudal máximo de diseño, así entonces:

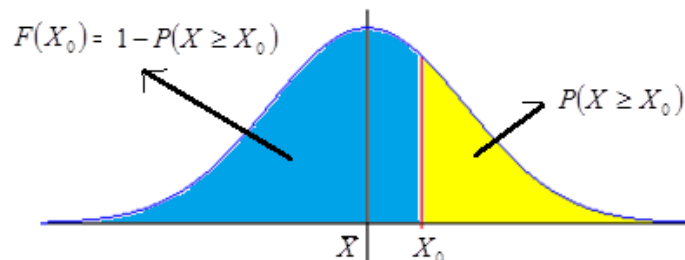
$$P(X \geq X_0) = \frac{1}{T} = \frac{1}{72.6} = 0.0138 = 1.38\%$$

**Caso de diseño para máximos**



Paso 4: Estimar la probabilidad acumulada asociada a  $X_0$

$$F(X_0) = P(X \leq X_0) = 1 - P(X \geq X_0) = 1 - 0.0138 = 0.986$$



Paso 5: Estimar los factores de frecuencia para cada una de las distribuciones de probabilidad usando las expresiones de cálculo del numeral 3:



**Distribución normal:** el factor de frecuencia corresponde a la variable tipificada  $z_T$  de la distribución normal estándar asociada a la probabilidad acumulada  $F(X)$ . Utilizando la tabla anexa de la distribución normal estándar.

$$F(X) = 0.986 \rightarrow z_T \approx 2.2$$

Por tanto para la distribución Normal

$$K_T = z_T = 2.2$$

El caudal para un periodo de retorno de 72.6 años, utilizando la distribución normal, está dado por:

$$X_{72.6} = \bar{X} + K_{Normal} S_x = 75.23 + 2.2(20) = 119.2 m^3 / s$$

El error estándar para la distribución se calcula como:

$$S_E = \frac{S_x}{\sqrt{N}} \left( 1 + \frac{z_T^2}{2} \right)^{0.5}$$

$$S_E = \frac{20}{\sqrt{18}} \left( 1 + \frac{2.2^2}{2} \right)^{0.5}$$

$$S_E = 8.71 m^3 / s$$

Finalmente las bandas de confianza para el caudal estimado mediante la distribución Normal serán:

$$Li = X_T - 1.96 S_E$$

Para el límite inferior:  $Li = 119.2 - 1.96 \times 8.71$

$$Li = 102.1 m^3 / s$$

$$Ls = X_T + 1.96 S_E$$

Para el límite superior:  $Ls = 119.2 + 1.96 \times 8.71$

$$Ls = 136.3 m^3 / s$$

Intervalo de 95% de confiabilidad: [102.1,136.3]

**Distribución Log-Normal:** el factor de frecuencia depende de la variable normal tipificada  $z_T$  y el coeficiente de variación  $C_{vx}$ , utilizando dicho factor no se requiere transformar los datos aplicando el logaritmo en base 10 de los datos; así pues:

$$K_T = \frac{\exp \left\{ z_T \left[ \ln(1 + C_{vx}^2) \right]^{0.5} - 0.5 \ln(1 + C_{vx}^2) \right\} - 1}{C_{vx}}$$

$$K_T = \frac{\exp \left\{ 2.2 \left[ \ln(1 + 0.266^2) \right]^{0.5} - 0.5 \ln(1 + 0.266^2) \right\} - 1}{0.266}$$

$$K_T = 2.69$$

Por tanto, El caudal para un periodo de retorno de 72.6 años, utilizando la distribución log-normal, está dado por:

$$X_{72.6} = \bar{X} + K_{LogNormal} S_x = 75.23 + 2.69(20) = 129.2 m^3 / s$$

El error estándar para la distribución log-normal se estima como:

$$S_E = \frac{S_x}{\sqrt{N}} \left[ 1 + (C_{vx}^3 + 3C_{vx}) K_T + (1/4)(C_{vx}^8 + 6C_{vx}^6 + 15C_{vx}^4 + 16C_{vx}^2 + 2) K_T^2 \right]^{0.5}$$

$$S_E = \frac{20}{\sqrt{18}} \left[ 1 + (0.27^3 + 3 \times 0.27) \times 2.69 + (1/4)(0.27^8 + 6 \times 0.27^6 + 15 \times 0.27^4 + 16 \times 0.27^2 + 2) \times 2.69^2 \right]^{0.5}$$

$$S_E = 14.22 m^3 / s$$

Las bandas de confianza para el caudal estimado mediante la distribución Normal serán:

$$Li = X_T - 1.96 S_E$$

Para el límite inferior:  $Li = 129.2 - 1.96 \times 14.22$

$$Li = 101.33 m^3 / s$$

$$Ls = X_T + 1.96 S_E$$

Para el límite superior:  $Ls = 129.2 + 1.96 \times 14.22$

$$Ls = 157.1 m^3 / s$$

Intervalo de 95% de confiabilidad: [101.33.157.1]

**Distribución Gumbel (GEV tipo I):** En dicha distribución el factor de frecuencia solo depende de frecuencia acumulada  $F(x)$ :

$$K_T = -0.45 - 0.7797 \ln\{-\ln[F(X)]\}$$

$$K_T = -0.45 - 0.7797 \ln\{-\ln[0.986]\}$$

$$K_T = 2.88$$

Por tanto, El caudal para un periodo de retorno de 72.6 años, utilizando la distribución Gumbel será:

$$X_{72.6} = X + K_{Gumbel} S_x = 75.23 + 2.88(20) = 132.9 m^3 / s$$

El error estándar para la distribución Gumbel se estima como:

$$S_E = \frac{S_x}{\sqrt{N}} (1 + 1.1396 K_T + 1.1 K_T^2)^{0.5}$$

$$S_E = \frac{20}{\sqrt{18}} (1 + 1.1396 \times 2.88 + 1.1 \times 2.88^2)^{0.5}$$

$$S_E = 17.26 m^3 / s$$

Las bandas de confianza para el caudal estimado mediante la distribución Gumbel serán:

$$Li = X_T - 1.96 S_E$$

Para el límite inferior:  $Li = 132.9 - 1.96 \times 17.26$

$$Li = 99.1 m^3 / s$$

$$Ls = X_T + 1.96 S_E$$

Para el límite superior:  $Ls = 132.9 + 1.96 \times 17.26$

$$Ls = 166.7 m^3 / s$$

Intervalo de 95% de confiabilidad: [99.1;166.7]

**Distribución Pearson Tipo III:** En dicha distribución el factor de frecuencia depende de la variable normal tipificada  $z_T$  y el coeficiente de asimetría  $g_x$

$$K_T = z_T + (z_T^2 - 1) \left( \frac{g_x}{6} \right) + \frac{1}{3} (z_T^3 - 6z_T) \left( \frac{g_x}{6} \right)^2 - (z_T^2 - 1) \left( \frac{g_x}{6} \right)^3 + (z_T) \left( \frac{g_x}{6} \right)^4 - \frac{1}{3} \left( \frac{g_x}{6} \right)^5$$

$$K_T = 2.2 + (2.2^2 - 1) \left( \frac{-0.4}{6} \right) + \frac{1}{3} (2.2^3 - 6 \times 2.2) \left( \frac{-0.4}{6} \right)^2 - (2.2^2 - 1) \left( \frac{-0.4}{6} \right)^3 + (2.2) \left( \frac{-0.4}{6} \right)^4 - \frac{1}{3} \left( \frac{-0.4}{6} \right)^5$$

$$K_T = 1.94$$

Por tanto, El caudal para un periodo de retorno de 72.6 años, utilizando la distribución Pearson tipo III, está dado por:

$$X_{72.6} = X + K_{Pearson} S_x = 75.23 + 1.94(20) = 114.03 m^3 / s$$

El error estándar para la distribución Pearson Tipo III se estima como:

$$S_E = \frac{S_x}{\sqrt{N}} \left[ 1 + g_x K_T + (K_T^2 / 2) (3g_x^2 / 4 + 1) + (3K_T)(W)(g_x + g_x^3 / 4) + 3(W^2)(2 + 3g_x^2 + 5g_x^4 / 8) \right]^{0.5}$$

Donde

$$W = \frac{(z_T^2 - 1)}{6} + \frac{4g_x(z_T^3 - 6z_T)}{6^3} - 3g_x^2 \frac{(z_T^2 - 1)}{6^3} + 4g_x^3 \frac{(z_T)}{6^4} - \frac{10g_x^4}{6^6}$$

$$W = \frac{(2.2^2 - 1)}{6} + \frac{4(-0.4)(2.2^3 - 6 \times 2.2)}{6^3} - 3(-0.4)^2 \frac{(2.2^2 - 1)}{6^3} + 4(-0.4)^3 \frac{(2.2)}{6^4} - \frac{10(-0.4)^4}{6^6}$$

$$W = 0.65$$

$$S_E = 8.3 m^3 / s$$

Las bandas de confianza para el caudal estimado mediante la distribución Gumbel serán:

$$Li = X_T - 1.96S_E$$

Para el límite inferior:  $Li = 114.03 - 1.96 \times 8.3$

$$Li = 97.8 m^3 / s$$

$$Ls = X_T + 1.96S_E$$

Para el límite superior:  $Ls = 114.03 + 1.96 \times 8.3$

$$Ls = 130.3 m^3 / s$$

Intervalo de 95% de confiabilidad: [97.8;130.3]

**Ejemplo 3.2:** Determinar los caudales máximos, con sus respectivas bandas de confianza para la quebrada La Concepción, para los periodos de retorno de 1.5, 2.33, 5, 10, 25, 50 y 100 años utilizando las distribuciones de probabilidad Normal, Lognormal, Gumbel y Pearson tipo III, Utilice los datos de caudales máximos mensuales de la estación Los Siprés.

**Solución:**

Inicialmente se deben estimar los principales momentos estadísticos de la muestra de caudales máximos anuales y su respectivo coeficiente de variación.

Media:  $\bar{X} = 75.23m^3 / s$

Desviación estándar:  $S_x = 20m^3 / s$

Coeficiente de asimetría:  $g_x = -0.41$

Coeficiente de variación:  $C_{vx} = 0.266$

Cantidad de datos:  $N = 18$

**Paso 1:** Calcular los factores de frecuencia para cada periodo de retorno

(1)	(2)	(3)	(4)	(5)	(6)	(7)
T	P	F(x)	z normal (K)	K lognormal	K Gumbel	k Pearson
1.5	0.67	0.33	-0.44**	-0.52	-0.52	-0.38
<b>2.33</b>	<b>0.43</b>	<b>0.57</b>	<b>0.18</b>	<b>0.05</b>	<b>0.00</b>	<b>0.24</b>
5	0.20	0.80	0.84	0.77	0.72	0.85
10	0.10	0.90	1.28	1.32	1.30	1.23
25	0.04	0.96	1.75	1.98	2.04	1.60
50	0.02	0.98	2.05	2.45	2.59	1.83
100	0.01	0.99	2.33	2.92	3.13	2.03

\*\* Cuando  $F(x) < 0.5$  se acude a las propiedades de simetría de la distribución normal.

**Columna (1):** Periodo de retorno

**Columna (2):** Probabilidad de ocurrencia que para eventos extremos máximos representa probabilidad de excedencia  $P$  estimada como:

$$P = \frac{1}{T}, \text{ utilizando la notación de columnas (2) = } \frac{1}{(1)}$$

Para el periodo de retorno de 2.33:  $P = \frac{1}{2.33} = 0.43$

**Columna (3):** representa la probabilidad de no excedencia  $F(x)$  estimada como

$F(x) = 1 - P$  Para caudales máximos. Utilizando la notación de columnas  
 (3) = 1 - (2).

Para el periodo de retorno de 2.33

$$F(x) = 1 - 0.43 = 0.57$$

**Columna (4):** Variable normal estandarizada asociada a una probabilidad de no excedencia  $F(x)$  estimada mediante el uso de la tabla de distribución normal. Dicha columna también representa el factor de frecuencia asociado a la distribución normal.

Para el periodo de retorno de 2.33  $F(x) = 0.57 \rightarrow z_T = 0.18$  usando la tabla.

**Columna (5):** Factor de frecuencia de la distribución log-normal. Estimado como:

$$K_T = \frac{\exp\left\{z_T \left[\ln(1 + C_{vx}^2)\right]^{0.5} - 0.5 \ln(1 + C_{vx}^2)\right\} - 1}{C_{vx}} \quad \text{o en la notación de columnas:}$$

$$(4) = \frac{\exp\left\{(3) \left[\ln(1 + C_{vx}^2)\right]^{0.5} - 0.5 \ln(1 + C_{vx}^2)\right\} - 1}{C_{vx}}$$

Para el periodo de retorno de 2.33:

$$K_{2.33} = \frac{\exp\left\{0.18 \left[\ln(1 + 0.266^2)\right]^{0.5} - 0.5 \ln(1 + 0.266^2)\right\} - 1}{0.266} = 0.05$$

**Columna (6):** Factor de frecuencia de la distribución Gumbel, estimado como:

$$K_T = -0.45 - 0.7797 \ln\{-\ln[F(X)]\} \quad \text{o en la notación de columnas:}$$

$$(6) = -0.45 - 0.7797 \ln\{-\ln[(3)]\}$$

Para el periodo de retorno de 2.33:

$$K_{2.33} = -0.45 - 0.7797 \ln\{-\ln[0.57]\} = 0.00$$

**Columna (7):** Factor de frecuencia de la distribución Pearson tipo III, estimado como:

$$K_T = z_T + (z_T^2 - 1) \left( \frac{g_x}{6} \right) + \frac{1}{3} (z_T^3 - 6z_T) \left( \frac{g_x}{6} \right)^2 - (z_T^2 - 1) \left( \frac{g_x}{6} \right)^3 + (z_T) \left( \frac{g_x}{6} \right)^4 - \frac{1}{3} \left( \frac{g_x}{6} \right)^5$$

$$(7) = z_T + ((4)^2 - 1) \left( \frac{g_x}{6} \right) + \frac{1}{3} ((7)^3 - 6(7)) \left( \frac{g_x}{6} \right)^2 - ((7)^2 - 1) \left( \frac{g_x}{6} \right)^3 + (7) \left( \frac{g_x}{6} \right)^4 - \frac{1}{3} \left( \frac{g_x}{6} \right)^5$$

$$K_{2.33} = 0.18 + (0.18^2 - 1) \left( \frac{g_x}{6} \right) + \frac{1}{3} (0.18^3 - 6 \times 0.18) \left( \frac{g_x}{6} \right)^2 - (0.18^2 - 1) \left( \frac{g_x}{6} \right)^3 + (0.18) \left( \frac{g_x}{6} \right)^4 - \frac{1}{3} \left( \frac{g_x}{6} \right)^5 = 0.24$$

**Paso 2:** Estimar el valor esperado de los caudales para los diferentes periodo de retorno utilizando

La ecuación básica para dicha estimación es la siguiente:

$$X_T = \bar{X} + K_T S_x$$

(1)	(8)	(9)	(10)	(11)
T	Q Normal	Q lognormal	Q gumbel	Q Pearson
1.5	66.43	64.80	64.76	67.59
<b>2.33</b>	<b>78.83</b>	<b>76.20</b>	<b>75.24</b>	<b>80.10</b>
5	92.03	90.55	89.60	92.29
10	100.83	101.59	101.30	99.80
25	110.24	114.87	116.07	107.29
50	116.24	124.24	127.03	111.79
100	121.84	133.68	137.92	115.81

**Columna (8):** Valor esperado de los caudales máximos para la distribución Normal.

$$X_T = \bar{X} + K_{TNormal} S_x \text{ o en notación de columnas (8) } = \bar{X} + (4) S_x$$

Para el periodo de retorno de 2.33

$$X_{2.33} = \bar{X} + 0.18 S_x = 75.23 + 0.18(20) = 80.1 m^3 / s$$

**Columna (9):** Valor esperado de los caudales máximos para la distribución Log-Normal.

$$X_T = \bar{X} + K_{TLog-Norm} S_x \text{ o en notación de columnas (9) } = \bar{X} + (5) S_x$$

Para el periodo de retorno de 2.33

$$X_{2.33} = \bar{X} + 0.18S_x = 75.23 + 0.05(20) = 76.2m^3 / s$$

**Columna (10):** Valor esperado de los caudales máximos para la distribución Gumbel.

$$X_T = \bar{X} + K_{Gumbel}S_x \text{ o en notación de columnas (10) } = \bar{X} + (6)S_x$$

Para el periodo de retorno de 2.33

$$X_{2.33} = \bar{X} + 0.0S_x = 75.23 + 0.0(20) = 75.23m^3 / s$$

**Columna (11):** Valor esperado de los caudales máximos para la distribución Pearson Tipo III.

$$X_T = \bar{X} + K_{Pearson}S_x \text{ o en notación de columnas (11) } = \bar{X} + (7)S_x$$

Para el periodo de retorno de 2.33

$$X_{2.33} = \bar{X} + 0.0S_x = 75.23 + 0.24(20) = 80.10m^3 / s$$

**Paso 3:** Estimación de los errores estándar para cada periodo de retorno y tipo de distribución:

(1)	(12)	(13)	(14)	(15)	(16)
T	Se Normal	Se lognormal	Se gumbel	W	Se Pearson
1.5	4.94	4.20	3.96	-0.15	5.45
<b>2.33</b>	<b>4.75</b>	<b>4.81</b>	<b>4.72</b>	<b>-0.15</b>	<b>5.06</b>
5	5.48	6.83	7.28	-0.02	4.90
10	6.36	8.78	9.84	0.15	5.34
25	7.50	11.32	13.26	0.38	6.83
50	8.30	13.18	15.87	0.55	8.38
100	9.09	15.08	18.49	0.74	10.21



**Columna (11):** Error estándar de la distribución Normal

$$S_E = \frac{S_x}{\sqrt{N}} \left( 1 + \frac{z_T^2}{2} \right)^{0.5} \text{ o en la notación de columnas:}$$

$$(11) = \frac{S_x}{\sqrt{N}} \left( 1 + \frac{(4)^2}{2} \right)^{0.5}$$

Para el periodo de retorno de 2.33

$$S_{E2.33} = \frac{S_x}{\sqrt{N}} \left( 1 + \frac{(0.18)^2}{2} \right)^{0.5} = \frac{20}{\sqrt{18}} \left( 1 + \frac{(0.18)^2}{2} \right)^{0.5} = 4.94 m^3 / s$$

**Columna (12):** Error estándar de la distribución log-normal

$$S_E = \frac{S_x}{\sqrt{N}} \left[ 1 + (C_{vx}^3 + 3C_{vx})K_T + (1/4)(C_{vx}^8 + 6C_{vx}^6 + 15C_{vx}^4 + 16C_{vx}^2 + 2)K_T^2 \right]^{0.5}$$

En notación de columnas

$$(12) = \frac{S_x}{\sqrt{N}} \left[ 1 + (C_{vx}^3 + 3C_{vx})(5) + (1/4)(C_{vx}^8 + 6C_{vx}^6 + 15C_{vx}^4 + 16C_{vx}^2 + 2)(5)^2 \right]^{0.5}$$

Para el periodo de retorno de 2.33

$$S_{E2.33} = \frac{S_x}{\sqrt{N}} \left[ 1 + (C_{vx}^3 + 3C_{vx})(0.05) + (1/4)(C_{vx}^8 + 6C_{vx}^6 + 15C_{vx}^4 + 16C_{vx}^2 + 2)(0.05)^2 \right]^{0.5}$$

$$S_{E2.33} = \frac{20}{\sqrt{18}} \left[ 1 + (0.27^3 + 3 \times 0.27)(0.05) + 0.25(0.27^8 + 6 \times 0.27^6 + 15 \times 0.27^4 + 16 \times 0.27^2 + 2)(0.05)^2 \right]^{0.5}$$

$$S_{E2.33} = 4.8 m^3 / s$$

**Columna (13):** Error estándar de la distribución Gumbel

$$S_E = \frac{S_x}{\sqrt{N}} \left( 1 + 1.1396K_T + 1.1K_T^2 \right)^{0.5} \text{ y en notación de columnas}$$

$$(13) = \frac{S_x}{\sqrt{N}} \left( 1 + 1.1396(6) + 1.1(6)^2 \right)^{0.5}$$

Para el periodo de retorno de 2.33

$$S_{E2.33} = \frac{20}{\sqrt{18}} \left( 1 + 1.1396(0.00) + 1.1(0.00)^2 \right)^{0.5} = 4.72 m^3 / s$$

**Columna (14):** Factor W para el error estándar de la distribución Pearson

$$W = \frac{(z_T^2 - 1)}{6} + \frac{4g_x(z_T^3 - 6z_T)}{6^3} - 3g_x^2 \frac{(z_T^2 - 1)}{6^3} + 4g_x^3 \frac{(z_T)}{6^4} - \frac{10g_x^4}{6^6}$$

En notación de columnas

$$(14) = \frac{((4)^2 - 1)}{6} + \frac{4g_x((4)^3 - 6z_T)}{6^3} - 3g_x^2 \frac{((4)^2 - 1)}{6^3} + 4g_x^3 \frac{((4))}{6^4} - \frac{10g_x^4}{6^6}$$

Para el periodo de retorno de 2.33

$$W_{2.33} = \frac{(0.18^2 - 1)}{6} + \frac{4(-0.41)(0.18^3 - 6z_T)}{6^3} - 3(-0.41)^2 \frac{(0.18^2 - 1)}{6^3} + 4(-0.41)^3 \frac{(0.18)}{6^4} - \frac{10(-0.41)^4}{6^6}$$

$$W_{2.33} = -0.15$$

**Columna (15):** Error estándar para la distribución Pearson Tipo III

$$S_E = \frac{S_x}{\sqrt{N}} \left[ 1 + g_x K_T + (K_T^2 / 2) (3g_x^2 / 4 + 1) + (3K_T)(W)(g_x + g_x^3 / 4) + 3(W^2) (2 + 3g_x^2 + 5g_x^4 / 8) \right]^{0.5}$$

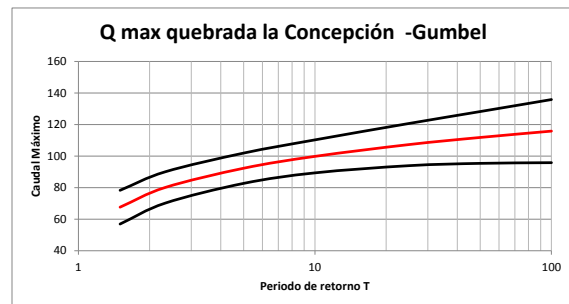
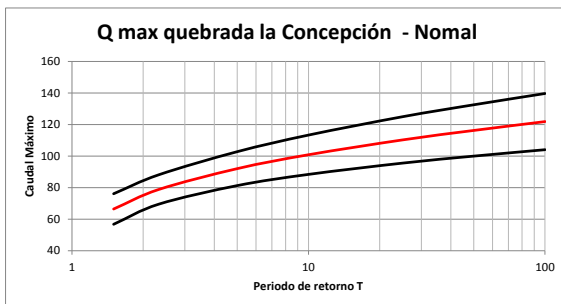
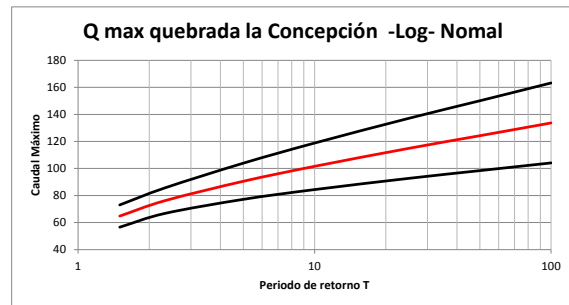
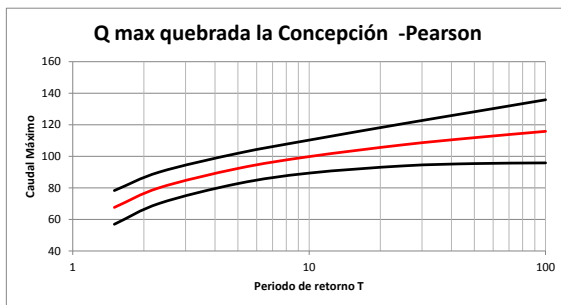
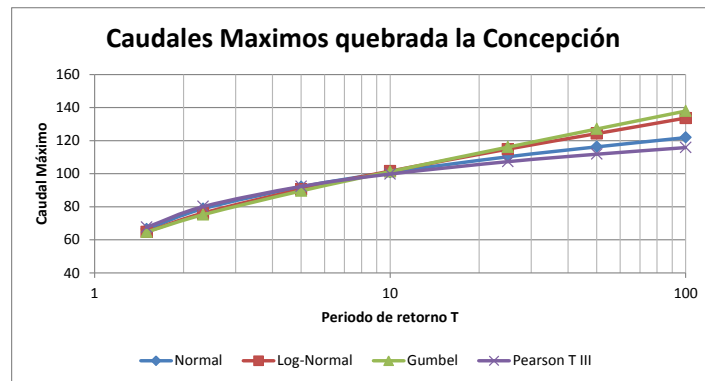
Utilizando la notación de columnas

$$S_E = \frac{S_x}{\sqrt{N}} \left[ 1 + g_x (7) + ((7)^2 / 2) (3g_x^2 / 4 + 1) + (3(7))(15)(g_x + g_x^3 / 4) + 3(15^2) (2 + 3g_x^2 + 5g_x^4 / 8) \right]^{0.5}$$

Para el periodo de retorno de 2.33

$$S_E = 5.06 m^3 / s$$

Distribución	Normal			Log-Normal			Gumbel			Pearson		
	Q	Li	Ls	Q	Li	Ls	Q	Li	Ls	Q	Li	Ls
T	Q	Li	Ls	Q	Li	Ls	Q	Li	Ls	Q	Li	Ls
1.5	66.4	56.7	76.1	64.8	56.6	73.0	64.8	57.0	72.5	67.6	56.9	78.3
2.33	78.8	69.5	88.1	76.2	66.8	85.6	75.2	66.0	84.5	80.1	70.2	90.0
5	92.0	81.3	102.8	90.6	77.2	103.9	89.6	75.3	103.9	92.3	82.7	101.9
10	100.8	88.4	113.3	101.6	84.4	118.8	101.3	82.0	120.6	99.8	89.3	110.3
25	110.2	95.5	124.9	114.9	92.7	137.1	116.1	90.1	142.1	107.3	93.9	120.7
50	116.2	100.0	132.5	124.2	98.4	150.1	127.0	95.9	158.1	111.8	95.4	128.2
100	121.8	104.0	139.7	133.7	104.1	163.2	137.9	101.7	174.2	115.8	95.8	135.8



Ejemplo 3.3: Se desea conocer el caudal mínimo con periodo de retorno de 10 años utilizando las distribuciones de probabilidad Normal, Lognormal, Gumbel y Pearson tipo III, con sus respectivos intervalos de confianza. Para ello considere los datos de Caudales mínimos mensuales de la estación Los Sipsrés (Figura 2).

VALORES MINIMOS MENSUALES DE CAUDALES (m3/seg)																	SISTEMA DE INFORMACION NACIONAL AMBIENTAL			
FECHA DE PROCESO : 2012/04/16										ESTACION : 23087180 SIRPES LOS							FECHA-INSTALACION 1973-MAY			
LATITUD 0622 N										DEPTO ANTIOQUIA							MUNICIPIO CONCEPCION		FECHA-SUSPENSION	
LONGITUD 7509 W										CORRIENTE QDA LA CONCEPCIO										
ELEVACION 1571 m.s.n.m																				
#O	EST	ENT	ENERO	FEBRE	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOST	SEPTI	OCTUB	NOVIE	DICIE	VR ANUAL					
1973	2	01	3.500	3.020	2.600	3.100	4.400	3.900	4.900	5.750	8.040	6.590	6.130	6.240	2.60					
1974	2	01	5.490	4.960	4.510	5.000	6.440	6.290	6.500	8.160	9.690	9.500	8.370	5.700	4.51					
1975	2	01	5.070	4.540	5.250	5.280	6.050	6.290	7.450	7.520	8.640	9.080	7.650	6.730	4.54					
1976	2	01	5.700	4.440	4.750	4.520	6.210	5.930	3.720	3.770	3.550	5.870	5.220	3.600	3.55					
1977	2	01	2.700	2.070	2.350	2.300	3.200	3.270	5.100	5.580	4.610	5.770	5.390	3.500	2.07					
1978	2	01	2.700	2.400	2.300	3.680	5.030	5.430	2.350	4.150	4.640	4.800	4.200	4.100	2.30					
1979	2	01	3.330	3.080	2.270	2.810	4.100	4.500	3.530	3.170	3.810	4.200	5.500	3.710	2.27					
1980	2	01	3.530	2.630	2.450	2.090	2.270	4.490	4.200	4.300	4.400	4.800	6.020	4.900	2.09					
1981	2	01	3.530	3.530	3.550	3.710	9.500	8.400	5.350	7.900	5.100	5.400	5.400	4.200	3.35					
1982	2	01	3.440	3.440	3.080	4.000	5.400	4.600	3.800	3.800	3.800	5.000	4.200	3.440	3.08					
1983	2	01	2.720	2.000	1.880	3.800	4.600	3.260	3.800	3.440	5.820	5.400	4.370	3.800	1.88					
1984	2	01	3.080	3.100	2.720	2.360	3.720	6.230	5.000	5.200	5.100	6.900	6.020	4.600	2.36					
1985	2	01	3.080	2.540	2.360	2.720	4.080	2.720	2.380	4.300	7.170	5.910	5.500	3.900	2.36					
1986	2	01	3.620	3.350	2.900	4.190	5.200	4.400	4.400	4.400	5.500	5.810	6.460	5.000	2.90					
1987	2	01	3.620	3.530	3.080	3.530	5.480	4.400	4.760	5.000	6.100	8.250	7.230	5.100	3.08					
1988	2	01	4.200	3.530	3.080	4.300	4.600	4.600	4.300	4.800	9.650	7.440	7.250	6.240	3.08					
1989	2	01	5.110	4.760	4.000	3.830	4.000	5.300	4.800	4.900	6.230	6.450	5.600	5.400	3.83					
1990	2	01	4.080	3.720	2.810	2.810	4.440	4.330	4.260	4.210	4.180	6.050	5.020	3.810	2.81					
MEDIOS			3.806	3.369	3.097	3.557	4.929	4.973	4.478	5.019	5.891	6.290	5.863	4.665	4.66					
MAXIMOS			5.700	4.960	5.250	5.280	9.500	8.400	7.450	8.160	9.690	9.500	8.370	6.730	9.69					
MINIMOS			2.700	2.000	1.880	2.090	2.270	2.720	2.350	3.170	3.550	4.200	4.200	3.440	1.88					

Figura 2: Caudales mínimos Qda la Concepción estación Sipres-Los

Paso 1: siguiendo las ecuaciones de los estimadores insesgados para los principales momentos estadísticos de una muestra aleatoria de datos se tiene que:

- Media:  $\bar{X} = 2.9m^3 / s$
- Desviación estándar:  $S_x = 0.79m^3 / s$
- Coefficiente de asimetría:  $g_x = 0.79$
- Coefficiente de variación:  $C_{vx} = 0.27$
- Cantidad de datos:  $N = 18$

Paso 2: por ser un estudio de caudales mínimos, la probabilidad de interés corresponde a la probabilidad de **no excedencia** de un evento de referencia  $X_0$  que corresponde al valor del caudal mínimos de interés, así entonces:

$$P(X \leq X_0) = \frac{1}{T} = \frac{1}{10} = 0.1 = 10\%$$



Paso 3: estimar la probabilidad acumulada asociada a  $X_0$

$$F(X_0) = P(X \leq X_0) = P = 0.1$$

Paso 4: Estimar los factores de frecuencia para cada una de las distribuciones de probabilidad usando las expresiones de cálculo del numeral 3.3:

**Distribución normal:** el factor de frecuencia corresponde a la variable tipificada  $z_T$  de la distribución normal estándar asociada a la probabilidad acumulada  $F(X)$ . Utilizando la tabla de la distribución normal estándar del anexo.

$$\text{Si } F(X) < 0.5 \rightarrow z_T < 0$$

$$\text{Luego } z_T = -z_0$$

Utilizando las propiedades de simetría de la distribución normal estándar.

$$F(z_T) = P(z \leq z_T) = P(z \leq -z_0) = P(z \geq z_0) = 1 - P(z \leq z_0) = 1 - 0.1 = 0.9$$

$$F(X) = 0.9 \rightarrow z_0 = 1.28$$

$$z_T = -z_0$$

$$z_T = -1.28$$

Por tanto para la distribución Normal

$$K_T = z_T = -1.28$$

El caudal mínimo con periodo de retorno de 10 años, utilizando la distribución normal, está dado por:

$$X_{10} = \bar{X} + K_{Normal} S_x = 2.9 + (-1.28)(0.79) = 1.91 m^3 / s$$

El error estándar para la distribución normal está dado por:

$$S_E = \frac{S_x}{\sqrt{N}} \left( 1 + \frac{z_T^2}{2} \right)^{0.5}$$

$$S_E = \frac{0.79}{\sqrt{18}} \left( 1 + \frac{(-1.28)^2}{2} \right)^{0.5}$$

$$S_E = 0.25 m^3 / s$$

Finalmente las bandas de confianza para el caudal estimado mediante la distribución Normal serán:

$$Li = X_T - 1.96S_E$$

Para el límite inferior:  $Li = 1.91 - 1.96 \times 0.25$

$$Li = 1.42 m^3 / s$$

$$Li = X_T + 1.96S_E$$

Para el límite superior:  $Li = 1.91 + 1.96 \times 0.25$

$$Li = 2.4 m^3 / s$$

Intervalo de 95% de confiabilidad: [2.41,3.39]

**Distribución Log-Normal:** el factor de frecuencia depende de la variable normal tipificada  $z_T$  y el coeficiente de variación  $C_{vx}$ , utilizando dicho factor no se requiere aplicar el logaritmo natural de los datos:

$$K_T = \frac{\exp \left\{ z_T \left[ \ln(1 + C_{vx}^2) \right]^{0.5} - 0.5 \ln(1 + C_{vx}^2) \right\} - 1}{C_{vx}}$$

$$K_T = \frac{\exp \left\{ (-1.28) \left[ \ln(1 + 0.27^2) \right]^{0.5} - 0.5 \ln(1 + 0.27^2) \right\} - 1}{0.27}$$

$$K_T = -1.16$$

Por tanto, El caudal mínimo para un periodo de retorno de 10 años, utilizando la distribución log-normal, está dado por:

$$X_{10} = \bar{X} + K_{LogNormal} S_x = 2.93 + (-1.16)(0.79) = 2.01 m^3 / s$$

El error estándar para la distribución log-normal está dado por:

$$S_E = \frac{S_x}{\sqrt{N}} \left[ 1 + (C_{vx}^3 + 3C_{vx}) K_T + (1/4)(C_{vx}^8 + 6C_{vx}^6 + 15C_{vx}^4 + 16C_{vx}^2 + 2) K_T^2 \right]^{0.5}$$

$$S_E = \frac{0.79}{\sqrt{18}} \left[ 1 + (0.27^3 + 3 \times 0.27) \times (-1.16) + (1/4)(0.27^8 + 6 \times 0.27^6 + 15 \times 0.27^4 + 16 \times 0.27^2 + 2) \times (-1.16)^2 \right]^{0.5}$$

$$S_E = 0.2 m^3 / s$$

las bandas de confianza para el caudal estimado mediante la distribución Normal serán:

$$Li = X_T - 1.96 S_E$$

Para el límite inferior:  $Li = 2.01 - 1.96 \times 0.2$

$$Li = 1.62 m^3 / s$$

$$Ls = X_T + 1.96 S_E$$

Para el límite superior:  $Ls = 2.01 + 1.96 \times 0.2$

$$Ls = 2.4 m^3 / s$$

Intervalo de 95% de confiabilidad: [1.62.2.4]

**Distribución Gumbel (GEV tipo I):** En dicha distribución el factor de frecuencia solo depende de frecuencia acumulada F(x):

$$K_T = -0.45 - 0.7797 \ln \{- \ln [F(X)]\}$$

$$K_T = -0.45 - 0.7797 \ln \{- \ln [0.1]\}$$

$$K_T = -1.1$$

Por tanto, El caudal mínimo para un periodo de retorno de 10 años, utilizando la distribución Gumbel, está dado por:

$$X_{10} = \bar{X} + K_{Gumbel} S_x = 2.9 + (-1.1)(0.79) = 2.05 m^3 / s$$

El error estándar para la distribución Gumbel se estima como:

$$S_E = \frac{S_x}{\sqrt{N}} \left(1 + 1.1396K_T + 1.1K_T^2\right)^{0.5}$$

$$S_E = \frac{0.79}{\sqrt{18}} \left(1 + 1.1396 \times (-1.1) + 1.1 \times (-1.1)^2\right)^{0.5}$$

$$S_E = 0.19m^3 / s$$

Las bandas de confianza para el caudal mínimo estimado mediante la distribución Gumbel serán:

$$Li = X_T - 1.96S_E$$

Para el límite inferior:  $Li = 2.05 - 1.96 \times 0.19$

$$Li = 1.68m^3 / s$$

$$Ls = X_T + 1.96S_E$$

Para el límite superior:  $Ls = 2.05 + 1.96 \times 0.19$

$$Ls = 2.42m^3 / s$$

Intervalo de 95% de confiabilidad: [1.68;2.42]

**Distribución Pearson Tipo III:** En dicha distribución el factor de frecuencia depende de la variable normal tipificada  $z_T$  y el coeficiente de asimetría  $g_x$

$$K_T = z_T + (z_T^2 - 1) \left(\frac{g_x}{6}\right) + \frac{1}{3} (z_T^3 - 6z_T) \left(\frac{g_x}{6}\right)^2 - (z_T^2 - 1) \left(\frac{g_x}{6}\right)^3 + (z_T) \left(\frac{g_x}{6}\right)^4 - \frac{1}{3} \left(\frac{g_x}{6}\right)^5$$

$$K_T = -1.17$$

Por tanto, El caudal mínimo para un periodo de retorno de 10 años, utilizando la distribución Pearson tipo III, está dado por:

$$X_{10} = \bar{X} + K_{Pearson} S_x = 2.9 + 1.94(0.79) = 2m^3 / s$$

El error estándar para la distribución Pearson Tipo III se estima como:

$$S_E = \frac{S_x}{\sqrt{N}} \left[1 + g_x K_T + (K_T^2 / 2) (3g_x^2 / 4 + 1) + (3K_T)(W)(g_x + g_x^3 / 4) + 3(W^2)(2 + 3g_x^2 + 5g_x^4 / 8)\right]^{0.5}$$

Donde



$$W = \frac{(z_T^2 - 1)}{6} + \frac{4g_x(z_T^3 - 6z_T)}{6^3} - 3g_x^2 \frac{(z_T^2 - 1)}{6^3} + 4g_x^3 \frac{(z_T)}{6^4} - \frac{10g_x^4}{6^6}$$

$$S_E = 0.18m^3 / s$$

Las bandas de confianza para el caudal estimado mediante la distribución Gumbel serán:

$$Li = X_T - 1.96S_E$$

Para el límite inferior:  $Li = 2 - 1.96 \times 0.18$

$$Li = 1.64m^3 / s$$

$$Ls = X_T + 1.96S_E$$

Para el límite superior:  $Ls = 2 + 1.96 \times 0.18$

$$Ls = 2.35m^3 / s$$

Intervalo de 95% de confiabilidad: [1.64;2.35]

**Ejemplo 3.4:** Determinar los caudales mínimos, con sus respectivas bandas de confianza para la quebrada La Concepción, para los periodos de retorno de 1.5, 2.33, 5, 10, 25, 50 y 100 años utilizando las distribuciones de probabilidad Normal, Lognormal, Gumbel y Pearson tipo III, Utilice los datos de caudales mínimos mensuales de la estación Los Siprés.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
TR	P	F(x)	z normal (K)	K lognormal	K gumbel	k Pearson
1.5	0.67	0.67	0.43	0.30	0.25	0.31
2.33	0.43	0.43	-0.18	-0.29	-0.32	-0.30
5	0.20	0.20	-0.84	-0.84	-0.82	-0.85
10	0.10	0.10	-1.28	-1.16	-1.10	-1.17
25	0.04	0.04	-1.75	-1.46	-1.36	-1.45
50	0.02	0.02	-2.05	-1.63	-1.51	-1.61
100	0.01	0.01	-2.33	-1.77	-1.64	-1.75

**Columna (1):** Periodo de retorno

**Columna (2):** Probabilidad de ocurrencia que para eventos extremos máximos representa probabilidad de excedencia  $P$  estimada como:

$$P = \frac{1}{T}, \text{ utilizando la notación de columnas } (2) = \frac{1}{(1)}$$

Para el periodo de retorno de 2.33:  $P = \frac{1}{2.33} = 0.43$

**Columna (3):** representa la probabilidad de no excedencia  $F(x)$  estimada como  $F(x) = P$  Para caudales mínimos

Utilizando la notación de columnas (3) = (2).

Para el periodo de retorno de 2.33

$$F(x) = P = 0.43$$

**Columna (4):** Variable normal estandarizada asociada a una probabilidad de no excedencia  $F(x)$  se estimada mediante el uso de la tabla de distribución normal. Dicha columna también representa el factor de frecuencia asociado a la distribución normal.

Para el periodo de retorno de 2.33  $F(x) = 0.43 \rightarrow z_T = -0.18$  usando la tabla y la propiedad de simetría de la distribución normal estándar.

**Columna (5):** Factor de frecuencia de la distribución log-normal. Estimado como:

$$K_T = \frac{\exp \left\{ z_T \left[ \ln(1 + C_{vx}^2) \right]^{0.5} - 0.5 \ln(1 + C_{vx}^2) \right\} - 1}{C_{vx}} \text{ o en la notación de columnas:}$$

$$(4) = \frac{\exp \left\{ (3) \left[ \ln(1 + C_{vx}^2) \right]^{0.5} - 0.5 \ln(1 + C_{vx}^2) \right\} - 1}{C_{vx}}$$

Para el periodo de retorno de 2.33:

$$K_{2.33} = \frac{\exp \left\{ (-0.18) \left[ \ln(1 + 0.27^2) \right]^{0.5} - 0.5 \ln(1 + 0.27^2) \right\} - 1}{0.27} = -0.29$$

**Columna (6):** Factor de frecuencia de la distribución Gumbel, estimado como:

$K_T = -0.45 - 0.7797 \ln\{-\ln[F(X)]\}$  o en la notación de columnas:

$$(6) = -0.45 - 0.7797 \ln\{-\ln[(3)]\}$$

Para el periodo de retorno de 2.33:

$$K_{2.33} = -0.45 - 0.7797 \ln\{-\ln[0.43]\} = -0.32$$

**Columna (7):** Factor de frecuencia de la distribución Pearson tipo III, estimado como:

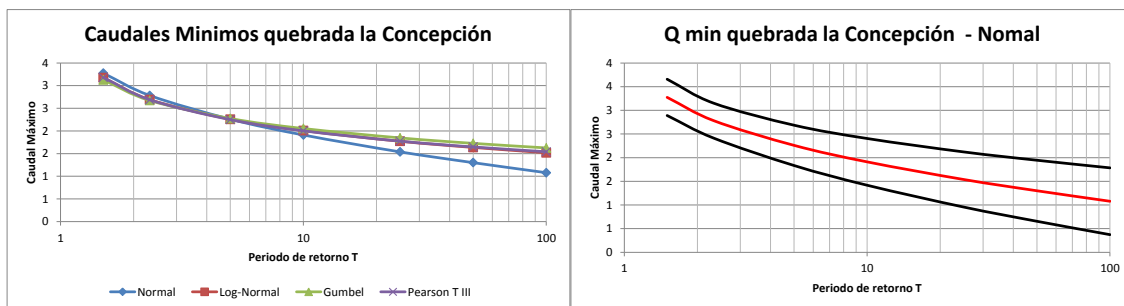
$$K_T = z_T + (z_T^2 - 1)\left(\frac{g_x}{6}\right) + \frac{1}{3}(z_T^3 - 6z_T)\left(\frac{g_x}{6}\right)^2 - (z_T^2 - 1)\left(\frac{g_x}{6}\right)^3 + (z_T)\left(\frac{g_x}{6}\right)^4 - \frac{1}{3}\left(\frac{g_x}{6}\right)^5$$

$$(7) = z_T + (4)^2 - 1\left(\frac{g_x}{6}\right) + \frac{1}{3}((7)^3 - 6(7))\left(\frac{g_x}{6}\right)^2 - ((7)^2 - 1)\left(\frac{g_x}{6}\right)^3 + (7)\left(\frac{g_x}{6}\right)^4 - \frac{1}{3}\left(\frac{g_x}{6}\right)^5$$

$$K_{2.33} = -0.30$$

Se continúa el procedimiento como en el ejercicio 3.2 para llegar a los siguientes resultados.

Distribucion	NORMAL			Log-Normal			Gumbel			Pearson		
	Q	Li	Ls	Q	Li	Ls	Q	Li	Ls	Q	Li	Ls
1.5	3.3	2.9	3.7	3.2	2.8	3.6	3.1	2.7	3.6	3.2	2.7	3.6
2.33	2.8	2.4	3.2	2.7	2.4	3.0	2.7	2.4	3.0	2.7	2.3	3.1
5	2.3	1.8	2.7	2.3	1.9	2.6	2.3	1.9	2.6	2.2	1.9	2.6
10	1.9	1.4	2.4	2.0	1.6	2.4	2.1	1.7	2.4	2.0	1.7	2.3
25	1.5	1.0	2.1	1.8	1.3	2.2	1.8	1.4	2.3	1.8	1.3	2.3
50	1.3	0.7	1.9	1.6	1.1	2.1	1.7	1.2	2.2	1.6	1.0	2.3
100	1.1	0.4	1.8	1.5	1.0	2.0	1.6	1.1	2.2	1.5	0.7	2.3



### 3.4 PRUEBAS DE BONDAD DE AJUSTE

Luego de analizar y aplicar las diferentes distribuciones de probabilidad a una muestra de datos hidrológicos para estimar eventos extremos asociados a diferentes periodos de retorno surge la pregunta de cuál es la mejor función de distribución de probabilidad para reproducir adecuadamente las leyes de probabilidad asociadas a la muestra de datos y obtener medidas más confiables de los caudales de diseño. Realmente no existe aún un consenso entre hidrólogos sobre cual distribución debería usarse. Es por ello que se vuelve necesaria la aplicación de pruebas de bondad de ajuste a fin de seleccionar la más adecuada fdp entre varias distribuciones.

Así pues, la bondad de ajuste de una distribución de probabilidad puede probarse comparando los valores teóricos y muestrales de las funciones de frecuencia relativa o acumulada de los datos. Una prueba adicional puede hacerse calculando la suma de los cuadrados de las diferencias entre los valores observados y los calculados, o el coeficiente de correlación de Pearson entre ambos datos.

#### 5.4.1 Prueba de bondad de ajuste $\chi^2$

El estadístico  $\chi^2$  estima el porcentaje de diferencia entre la función de frecuencia relativa asociada a la muestra de datos y la distribución de probabilidades teórica. El valor muestra del histograma de frecuencias asociada al intervalo  $i$  (frecuencia relativa) será  $f(x_i) = n_i/n$  donde  $n_i$  representa el número de ocurrencias observadas en el intervalo  $i$  y  $n$  corresponde al número total de datos presentes en la muestra. El valor teórico de la distribución de probabilidades está dado por  $p(x_i) = F(x_i) - F(x_{i-1})$ . El estadístico de la prueba  $\chi^2$  está dado por

$$\chi^2 = \sum_{i=1}^m \frac{n[f(x_i) - p(x_i)]^2}{p(x_i)}$$

Donde  $m$  representa el número de intervalos del histograma de frecuencias. La función de distribución de probabilidades  $\chi^2$  se encuentra tabulada en muchos textos de estadística, y depende del número  $\nu$  de grados de libertad y es igual a

$$\nu = m - p - 1$$

Siendo  $p$  el número de parámetros que posee la función de distribución de probabilidades que se desea ajustar a la muestra de datos, además se debe seleccionar un nivel de significancia  $\alpha$  para seleccionar el valor de  $\chi^2$



## BONDAD DE AJUSTE A LA DISTRIBUCION NORMAL

Previo al análisis se deben estimar los caudales máximos anuales a fin de construir la muestra de caudales para el análisis de bondad de ajuste. Además se requiere estimar los principales momentos estadísticos de la muestra de datos, para el caso de los caudales máximos anuales del Rio Negro en la estación Puente Real se tiene que:

Media:  $\bar{X} = 46.35m^3 / s$

Desviación estándar:  $S_x = 21.34m^3 / s$

Coefficiente de asimetría:  $g_x = 1.45m^3 / s$

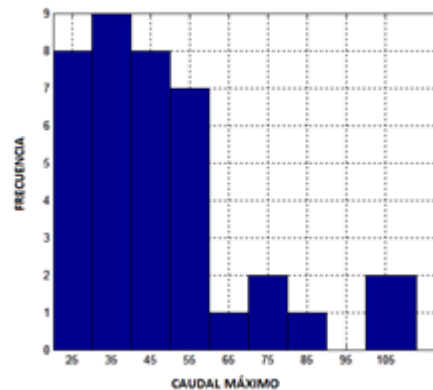
Coefficiente de variación:  $C_{vx} = 0.454$

Cantidad de datos:  $N = 38$

Paso 1: Construir el histograma de frecuencias de los datos, para ello se identifican los valores máximo y mínimo de la muestra y define el ancho y numero de intervalos del histograma. Para los 38 datos de caudales máximos ( $n = 38$ ) de la estación Pte. Real el valor máximo de caudal es de  $113 m^3/s$  y el valor mínimo es de 21, lo que permite definir por comodidad 9 intervalos que se presentan en la Figura ..., para cada uno de los intervalos se identifica el número de datos ubicado en el rango asociado ( $n_i$ ).

INTERVALO	RANGO	ni
1	<30	8
2	30-40	9
3	40-50	8
4	50-60	7
5	60-70	1
6	70-80	2
7	80-90	1
8	90-100	0
9	>100	2

HISTOGRAMA DE FRECUENCIAS DE CAUDALES MÁXIMOS PARA LA ESTACIÓN PTE. REAL SOBRE EL RÍO NEGRO



Paso 2: Estimar la frecuencia relativa asociada a cada intervalo  $f(x_i) = n_i/n$  donde  $n_i$  representa el número de ocurrencias observadas en el intervalo  $i$  y  $n$  corresponde al número total de datos presentes en la muestra. Para el presente ejemplo, con  $n = 38$   $f(x_i)$  Representa el valor empírico de función de distribución de probabilidades.

INTERVALO	RANGO	$n_i$	$f(x_i) = n_i/n$
1	< 30	8	0.211
2	30-40	9	0.237
3	40-50	8	0.211
4	50-60	7	0.184
5	60-70	1	0.026
6	70-80	2	0.053
7	80-90	1	0.026
8	90-100	0	0.000
9	> 100	2	0.053

Paso 3: Estimar la frecuencia empírica acumulada  $F_E(x_i)$  asociada a cada intervalo: corresponde a la sumatoria de la frecuencias relativas hasta el intervalo  $i$ .

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i) = \sum_{k=1}^i f(x_k)$
1	< 30	8	0.211	0.211
2	30-40	9	0.237	0.447
3	40-50	8	0.211	0.658
4	50-60	7	0.184	0.842
5	60-70	1	0.026	0.868
6	70-80	2	0.053	0.921
7	80-90	1	0.026	0.947
8	90-100	0	0.000	0.947
9	> 100	2	0.053	1.000

Paso 4: Estimar la variable normal estandarizada asociada al límite superior de cada rango ( $LSR$ ):

Para cada caso:  $z_i = \frac{LSR - \bar{X}}{S_x}$  Por ejemplo, para el intervalo 1

$$z_1 = \frac{LSR - \bar{X}}{S_x} = \frac{30 - 46.35}{21.34} = -0.78$$

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i) = \sum_{k=1}^i f(x_k)$	$LSR$	$z$
1	< 30	8	0.211	0.211	30	-0.78
2	30-40	9	0.237	0.447	40	-0.30
3	40-50	8	0.211	0.658	50	0.17
4	50-60	7	0.184	0.842	60	0.65
5	60-70	1	0.026	0.868	70	1.12
6	70-80	2	0.053	0.921	80	1.60
7	80-90	1	0.026	0.947	90	2.07
8	90-100	0	0.000	0.947	100	2.55
9	> 100	2	0.053	1.000		

Paso 5: Para cada  $z_i$  obtenido en el paso anterior se estima la probabilidad acumulada teórica  $F(z_i)$  utilizando la tabla de la distribución Normal o cualquier otra aproximación, por ejemplo, Excel posee la función DISTR.NORM.ESTAND( $z$ ) para estimar la probabilidad acumulada de  $z_i$  usando la distribución normal estándar.

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i)$	$LSR$	$z$	$F(z_i)$
1	< 30	8	0.211	0.211	30	-0.78	0.22
2	30-40	9	0.237	0.447	40	-0.30	0.38
3	40-50	8	0.211	0.658	50	0.17	0.57
4	50-60	7	0.184	0.842	60	0.65	0.74
5	60-70	1	0.026	0.868	70	1.12	0.87
6	70-80	2	0.053	0.921	80	1.60	0.95
7	80-90	1	0.026	0.947	90	2.07	0.98
8	90-100	0	0.000	0.947	100	2.55	0.99
9	> 100	2	0.053	1.000			1.00

Paso 6: Para cada intervalo se debe estimar la probabilidad teórica  $p(x_i)$  donde:

$$p(x_i) = F(x_i) - F(x_{i-1})$$

Por ejemplo, para  $i=1$ :

$$p(x_1) = P(x \leq 30) = P\left(z \leq \frac{30 - 46.35}{21.34}\right) = P(z \leq -0.78) = F(-0.78) = 0.22$$

Para  $i=2$



$$p(x_2) = P(30 \leq x < 40) = P(x \leq 40) - P(x \leq 30) = F(-0.78) - F(-0.78) = 0.38 - 0.22 = 0.16$$

y así sucesivamente

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i)$	$LSR$	$z$	$F(z_i)$	$p(x_i)$
1	< 30	8	0.211	0.211	30	-0.78	0.22	0.22
2	30-40	9	0.237	0.447	40	-0.30	0.38	0.16
3	40-50	8	0.211	0.658	50	0.17	0.57	0.19
4	50-60	7	0.184	0.842	60	0.65	0.74	0.17
5	60-70	1	0.026	0.868	70	1.12	0.87	0.13
6	70-80	2	0.053	0.921	80	1.60	0.95	0.08
7	80-90	1	0.026	0.947	90	2.07	0.98	0.04
8	90-100	0	0.000	0.947	100	2.55	0.99	0.01
9	> 100	2	0.053	1.000			1.00	0.01

Paso 7 : Estimar el estadístico  $\chi^2$  como:

$$\chi^2 = \sum_{i=1}^m \frac{n[f(x_i) - p(x_i)]^2}{p(x_i)}$$

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i)$	$LSR$	$z$	$F(z_i)$	$p(x_i)$	$\chi^2$
1	< 30	8	0.211	0.211	30	-0.78	0.22	0.22	0.01
2	30-40	9	0.237	0.447	40	-0.30	0.38	0.16	1.28
3	40-50	8	0.211	0.658	50	0.17	0.57	0.19	0.11
4	50-60	7	0.184	0.842	60	0.65	0.74	0.17	0.03
5	60-70	1	0.026	0.868	70	1.12	0.87	0.13	3.06
6	70-80	2	0.053	0.921	80	1.60	0.95	0.08	0.27
7	80-90	1	0.026	0.947	90	2.07	0.98	0.04	0.10
8	90-100	0	0.000	0.947	100	2.55	0.99	0.01	0.52
9	> 100	2	0.053	1.000			1.00	0.01	15.73
<b>Suma</b>									<b>21.10</b>

$$\chi^2 = \sum_{i=1}^m \frac{n[f(x_i) - p(x_i)]^2}{p(x_i)} = 21.1$$

Paso 8 : Estimar el valor crítico del estadístico  $\chi^2$ ,  $\chi_{Crítico}^2$  utilizando un nivel de significancia de 0.05 y el número de grados de libertad asociado al tipo de distribución:

El número de grados de libertad está dado por:

$$v = m - p - 1$$

Siendo  $P$  el número de parámetros que posee la función de distribución de probabilidades, que para el caso de la distribución normal solo posee dos parámetros (la media y la desviación estándar), y  $m$  el número de intervalos que para el presente ejemplo corresponde a 9.

$$v = 9 - 2 - 1 = 6$$

Utilizando la tabla de la distribución  $\chi^2$ ,  $\chi_{Crítico}^2 = \chi_{v,\alpha}^2 = \chi_{6,0.05}^2 = 12.59$

Finalmente para probar que los datos se ajustan a la distribución Normal el valor  $\chi^2$  estimado mediante la muestra de datos debe ser inferior al  $\chi_{Crítico}^2$ , para el presente caso  $21.10 \geq 12.59$ , por tanto los datos *No se ajustan a la distribución normal*.

Comentarios:

- Para el análisis de bondad de ajuste de las diversas distribuciones de frecuencia siempre siguen los pasos 1,2,3, 6, 7, 8 de la misma forma que en el presente ejemplo de la distribución normal.
- El paso 4 es un paso intermedio necesario para efectuar los cálculos en la distribución normal
- El paso 5 depende del tipo de función teórica de probabilidades y sus parámetros.

#### 5.4.1.2 Bondad de ajuste para la distribución Log-Normal

Probar la bondad de ajuste de los datos a la distribución log normal significa que una nueva variable, que equivale al logaritmo en base 10 de los datos de caudales máximos se distribuye siguiendo la fdp Normal. Es por ello que inicialmente se deben transformar los datos estimando su logaritmo en base 10 y luego estimarse los principales momentos estadísticos de los datos.

Media:  $\bar{X} = 1.63$   
 Desviación estándar:  $S_x = 0.18$   
 Coeficiente de asimetría:  $g_x = 0.36$   
 Coeficiente de variación:  $C_{vx} = 0.11$   
 Cantidad de datos:  $N = 38$

Para el caso de la distribución Log-Normal se siguen los pasos del 1 al 8 del ejemplo anterior pero con los datos transformados en el espacio logarítmico:

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i)$	$LSR$	$z$	$F(z_i)$	$p(x_i)$	$\chi^2$
1	< 1.4	4	0.105	0.105	1.4	-1.27	0.10	0.10	0.0037
2	1.4-1.5	5	0.132	0.237	1.5	-0.71	0.24	0.14	0.0046
3	1.5-1.6	8	0.211	0.447	1.6	-0.16	0.44	0.20	0.0224
4	1.6-1.7	8	0.211	0.658	1.7	0.40	0.65	0.22	0.0084
5	1.7-1.8	8	0.211	0.868	1.8	0.95	0.83	0.18	0.2702
6	1.8-1.9	2	0.053	0.921	1.9	1.51	0.93	0.10	0.9769
7	1.9-2.0	1	0.026	0.947	2	2.07	0.98	0.05	0.3215
8	> 2.0	2	0.053	1.000			1.00	0.02	2.1680
suma									3.78

En este caso  $\nu = 8 - 2 - 1 = 5$ . Utilizando la tabla de la distribución  $\chi^2$ :

$$\chi_{\text{Crítico}}^2 = \chi_{\nu, \alpha}^2 = \chi_{5, 0.05}^2 = 11.07$$

Para probar que los datos se ajustan a la distribución Log-Normal el valor  $\chi^2$  estimado mediante la muestra de datos debe ser inferior al  $\chi_{\text{Crítico}}^2$ , para el presente caso  $3.78 \geq 11.07$ , por tanto los datos *Si se ajustan a la distribución Log-normal*.

## PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN GUMBEL

Con respecto a la distribución normal solo cambia el paso 5 para la estimación de la función de distribución de probabilidades acumulada teórica. Para este caso se utilizan las ecuaciones de la distribución Gumbel.

Para el caso de los caudales máximos anuales del Rio Negro en la estación Puente Real se tiene que:

Media:  $\bar{X} = 46.35m^3 / s$   
 Desviación estándar:  $S_x = 21.34m^3 / s$   
 Coeficiente de asimetría:  $g_x = 1.45m^3 / s$   
 Coeficiente de variación:  $C_{vx} = 0.454$   
 Cantidad de datos:  $N = 38$

Para la distribución Gumbel la función de probabilidades acumulada está dada por:

$$F(c) = \exp[-\exp(-c)]$$

Donde

$$c = a(x - b)$$

Siendo

$$a = \frac{\sigma_Y}{S_x} \quad b = \bar{X} - \frac{\mu_Y}{S_x}$$

Donde  $\sigma_Y$ ,  $\mu_Y$  dependen de la cantidad de datos presentes en la muestra y se puede consultar en tablas o mediante el uso de algunas ecuaciones de ajuste como las estimadas por Rojo (2013):

$$\mu_Y = 0.0262 \ln(N) + 0.4436 = 0.0262 \ln(38) + 0.4436 = 0.54$$

$$\sigma_Y = 0.133 \ln(N) + 0.7467 = 0.133 \ln(38) + 0.7467 = 1.23$$

Luego:  $a = \frac{\sigma_Y}{S_x} = \frac{1.23}{21.34} = 0.0585$   $b = \bar{X} - \frac{\mu_Y}{a} = 46.35 - \frac{0.54}{0.0585} = 37.14$

En este caso el paso 4 de ejemplos anteriores se reemplaza por el cálculo de la variable  $c$  para el límite superior de cada intervalo.

Para  $i=1$ :  $c = a(LSR_1 - b) = 0.0585(30 - 37.14) = -0.42$

Para  $i=2$ :  $c = a(LSR_2 - b) = 0.0585(40 - 37.14) = 0.17$

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i) = \sum_{k=1}^i f(x_k)$	$LSR$	$c$
1	< 30	8	0.211	0.211	30	-0.42
2	30-40	9	0.237	0.447	40	0.17
3	40-50	8	0.211	0.658	50	0.75
4	50-60	7	0.184	0.842	60	1.34
5	60-70	1	0.026	0.868	70	1.92
6	70-80	2	0.053	0.921	80	2.51
7	80-90	1	0.026	0.947	90	3.09
8	90-100	0	0.000	0.947	100	3.68
9	> 100	2	0.053	1.000		

Finalmente para cada  $c_i$  se estima el valor de probabilidad acumulada  $F(c)$  como:

$F(c) = \exp[-\exp(-c)]$  y se continua con los pasos sugeridos ejemplos anteriores.

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i)$	$LSR$	$c$	$F(c_i)$	$p(x_i)$	$\chi^2$	
1	< 30	8	0.211	0.211	30	-0.42	0.22	0.22	0.0128	
2	30-40	9	0.237	0.447	40	0.17	0.43	0.21	0.1299	
3	40-50	8	0.211	0.658	50	0.75	0.62	0.20	0.0470	
4	50-60	7	0.184	0.842	60	1.34	0.77	0.14	0.4061	
5	60-70	1	0.026	0.868	70	1.92	0.86	0.09	1.8815	
6	70-80	2	0.053	0.921	80	2.51	0.92	0.06	0.0178	
7	80-90	1	0.026	0.947	90	3.09	0.96	0.03	0.0643	
8	90-100	0	0.000	0.947	100	3.68	0.98	0.02	0.7380	
9	> 100	2	0.053	1.000			1.00	0.02	1.1607	
									suma	4.46

En este caso  $\nu = 9 - 2 - 1 = 6$ . Utilizando la tabla de la distribución  $\chi^2$ :

$$\chi_{\text{Crítico}}^2 = \chi_{\nu, \alpha}^2 = \chi_{6, 0.05}^2 = 12.59$$

Para probar que los datos se ajustan a la distribución Log-Normal el valor  $\chi^2$  estimado mediante la muestra de datos debe ser inferior al  $\chi_{\text{Crítico}}^2$ , para el presente caso  $44.6 < 12.59$ , por tanto los datos *Si se ajustan a la distribución Gumbel*.

## Caso de la distribución Log-Gumbel

Probar la bondad de ajuste de los datos a la distribución log Gumbel significa que una nueva variable, que equivale al logaritmo en base 10 de los datos de caudales máximos se distribuye siguiendo la fdp Gumbel. Es por ello que inicialmente se deben transformar los datos estimando su logaritmo en base 10 y luego estimarse los principales momentos estadísticos de los datos.

Media:  $\bar{X} = 1.63$   
 Desviación estándar:  $S_x = 0.18$   
 Coeficiente de asimetría:  $g_x = 0.36$   
 Coeficiente de variación:  $C_{vx} = 0.11$   
 Cantidad de datos:  $N = 38$

$$\mu_Y = 0.0262 \ln(N) + 0.4436 = 0.0262 \ln(38) + 0.4436 = 0.54$$

$$\sigma_Y = 0.133 \ln(N) + 0.7467 = 0.133 \ln(38) + 0.7467 = 1.23$$

Luego:  $a = \frac{\sigma_Y}{S_x} = \frac{1.23}{0.18} = 6.84$        $b = \bar{X} - \frac{\mu_Y}{a} = 1.63 - \frac{0.54}{6.84} = 1.55$

Como en el caso anterior se procede al cálculo de la variable  $c$  para el límite superior de cada intervalo.

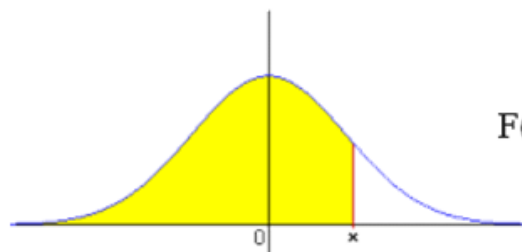
Para  $i=1$ :  $c = a(LSR_1 - b) = 6.84(1.4 - 1.55) = -1.02$

Para  $i=2$ :  $c = a(LSR_2 - b) = 6.84(1.5 - 1.55) = -0.34$

INTERVALO $i$	RANGO	$n_i$	$f(x_i) = n_i/n$	$F_E(x_i)$	$LSR$	$c$	$F(c_i)$	$p(x_i)$	$\chi^2$
1	< 1.4	4	0.105	0.105	1.4	-1.02	0.06	0.06	1.1560
2	1.4-1.5	5	0.132	0.237	1.5	-0.34	0.25	0.18	0.5637
3	1.5-1.6	8	0.211	0.447	1.6	0.35	0.49	0.25	0.2034
4	1.6-1.7	8	0.211	0.658	1.7	1.03	0.70	0.21	0.0022
5	1.7-1.8	8	0.211	0.868	1.8	1.71	0.84	0.14	1.5793
6	1.8-1.9	2	0.053	0.921	1.9	2.40	0.91	0.08	0.3137
7	1.9-2.0	1	0.026	0.947	2	3.08	0.96	0.04	0.2244
8	> 2.0	2	0.053	1.000			1.00	0.04	0.0516
								suma	4.09

## Tabla distribución Normal

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
<b>0,0</b>	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
<b>0,1</b>	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
<b>0,2</b>	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
<b>0,3</b>	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
<b>0,4</b>	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
<b>0,5</b>	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
<b>0,6</b>	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
<b>0,7</b>	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
<b>0,8</b>	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
<b>0,9</b>	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
<b>1,0</b>	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
<b>1,1</b>	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
<b>1,2</b>	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
<b>1,3</b>	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
<b>1,4</b>	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
<b>1,5</b>	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
<b>1,6</b>	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
<b>1,7</b>	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
<b>1,8</b>	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
<b>1,9</b>	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
<b>2,0</b>	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
<b>2,1</b>	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
<b>2,2</b>	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
<b>2,3</b>	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
<b>2,4</b>	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
<b>2,5</b>	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
<b>2,6</b>	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
<b>2,7</b>	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
<b>2,8</b>	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
<b>2,9</b>	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
<b>3,0</b>	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990



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