



## 2. Construcción de las curvas Intensidad-Duración-Periodo de retorno

### 2.1 Obtención de datos

#### 2.1.1 Instrumentos de medición de la precipitación

##### Pluviómetro

El *pluviómetro* es un instrumento que se emplea en las estaciones meteorológicas para medir la altura de la precipitación en un rango de tiempo. Las unidades empleadas en esta medida son milímetros de altura. El diseño básico de un pluviómetro consiste en un recipiente de entrada, llamado balancín, por donde el agua ingresa a través de un embudo hacia un colector donde el agua se acumula y puede medirse visualmente con una regla graduada o mediante el peso del agua contenida en el depósito. Asimismo, el balancín oscila a volumen constante de agua acumulada durante la precipitación, permitiendo el registro mecánico o eléctrico de la intensidad de lluvia precipitada.

Existen dos modelos básicos de pluviómetros: de lectura directa y registradores. Los de lectura directa tienen un recipiente y un embudo. Cada 24 horas (de 8:00 A.M a 8:00 A.M del día siguiente) se vacía el recipiente en una probeta graduada con una sección diez veces menor que la de recepción, con lo que es posible establecer una relación entre la altura en la probeta y la precipitación en milímetros por metro cuadrado.

Los pluviómetros registradores pueden ser de tres tipos: de pesada, de cuba basculante o de flotador, según el procedimiento que empleen para registrar la medición una vez alcanzado cierto nivel.

Normalmente la lectura se realiza cada 24 horas. Un litro caído en un metro cuadrado alcanzaría una altura de un milímetro. Para la medida de nieve se considera que el espesor de nieve equivale aproximadamente a diez veces el espesor de agua.





## Pluviógrafo

Este es un instrumento que sirve para registrar en forma continua la cantidad total y la duración de lluvia caída en milímetros(mm), de los registros puede definirse no sólo la altura de la precipitación caída sino también, cuanto ha caído, permitiendo analizar la distribución de la lluvia en el tiempo. El pluviógrafo que se utiliza normalmente en las estaciones es de sistema Hellman de Sifón.

El receptor del pluviógrafo es análogo al pluviómetro Hellman pero este va unido a una caja cilíndrica de mayor diámetro y de una altura de unos 110 cms, en la que se aloja debidamente protegido el sistema del aparato, y un depósito colector.

El agua recogida en el receptor pasa por un embudo y un tubo, el mecanismo de registro, constituido por un cilindro en cuyo interior hay un flotador que se desplaza verticalmente y esta conectado a un brazo de palanca con una pluma que registra en la banda, colocada sobre un tambor con un sistema de relojería la precipitación captada. El sistema de descarga del cilindro en que se aloja el flotador es de sifón.

La instalación del pluviógrafo debe guardar las mismas precauciones que las del pluviómetro tratando de que el agua captada represente lo mejor posible el volumen precipitable en el área circundante. El pluviógrafo deberá estar especialmente protegido de los efectos del viento.

La altura de la boca del pluviógrafo sera de 1.50 m, sobre el suelo, y su superficie quedará perfectamente horizontal, es muy importante la nivelación del aparato, para que su funcionamiento sea correcto.

Las bandas o gráficas que se ajustan al tambor, pueden ser diaria, semanal o mensual. Las diarias se usan más en periodos o zonas lluviosas, la semanal



**Figura 2.1.** Pluviógrafo





en lugares donde la lluvia no es diaria y las mensuales en períodos de estación seca o verano.

Su lectura se hará especialmente a las 8:00 A.M; la utilización del pluviógrafo es importante porque determina la intensidad de las precipitaciones, que es el factor fundamental para la clasificación de la precipitación en débil, moderado o fuerte.

### 2.1.2 CURVA MASA

La curva masa es la variación de la lluvia en un periodo de 24 hrs. Para obtenerla es necesario contar con el registro pluviográfico, como el mostrado en la figura 2.2.

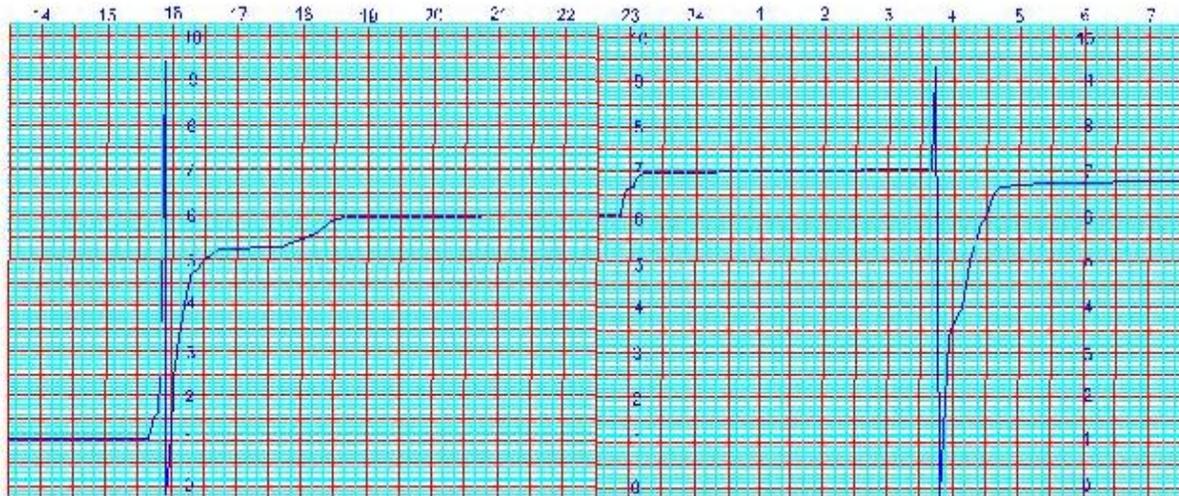


Figura 2.2. Registro generado por el pluviógrafo.

Si se eliminan del registro las líneas verticales descendentes y se grafica de forma continua y acumulada la altura de precipitación se obtendrá una curva que se conoce como *curva masa* de precipitación (figura 2.3) caracterizándose por formarse de curvas ascendentes y las líneas horizontales.

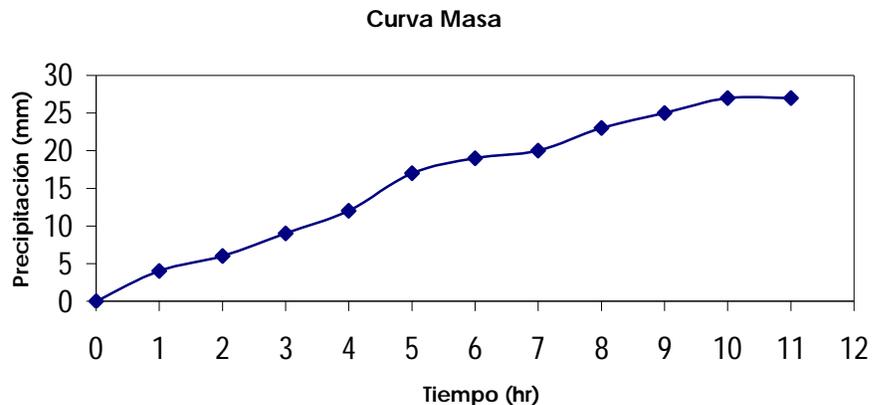
Es posible graficar la precipitación en forma de un diagrama de barras siendo conocida esta representación como hietograma, para su elaboración se debe establecer un intervalo de tiempo constante  $\Delta t$  y obtener la altura de precipitación para ese incremento de tiempo, estos





valores se graficarán en el eje de las "Y" mientras que en el eje de las "X" se ubicarán los incrementos de tiempo. Esta representación se conoce como hietograma de alturas de precipitación.

Si ahora la precipitación se divide entre el incremento de tiempo se obtendrá la intensidad de la precipitación siendo sus unidades en *mm/hr*. Esta representación gráfica se conocerá como hietograma de intensidades.



**Figura 2.3.** Curva Masa

## 2.2 Métodos de Regresión

### 2.2.1 Método de regresión lineal simple

Las tecnicas de regresión son el medio para estimar los parametros de un modelo matemático que expresa la relación de una variable dependiente o respuesta Y, la cual no se puede controlar en un experimento, en función de una o más variables independientes o de regresión  $X_1, X_2, \dots, X_k$ , las cuales se miden con un error despreciable y en algunos casos se controlan en el experimeto.

La relación fija para un conjunto de datos experimentales se caracteriza por una ecuación de predicción que recibe el nombre de ecuación de regresión. Para el caso de una variable dependiente Y, y una sola variable independiente X, se puede plantear un modelo de regresión lineal simple porblación del tipo.





$$\mu_{Y|X} = \alpha + \beta x \quad 2.1$$

Donde

$\alpha, \beta$  parámetros que se estiman a partir de los datos muestrales  
 $\mu, x$  valor esperado de Y dado que X ha ocurrido.

Si se considera como estimadores de  $\mu_{Y|X}, \alpha$  y  $\beta$ ,  $\hat{y}$ ,  $a$  y  $b$ , entonces la línea de regresión ajustada se puede plantear como:

$$\hat{y} = a + bx \quad 2.2$$

El símbolo  $\hat{y}$  se utiliza para distinguir entre el valor estimado que da la línea de regresión muestral y un valor experimental real observado  $y_i$  para algún valor de  $x_i$ .

El modelo del tipo de expresión (2.2) no es exacto físicamente, puesto que X y Y, son generalmente variables aleatorias y, mas aún, su dependencia puede no conocerse en forma exacta.

Por lo anterior es mas preciso escribir el modelo como:

$$Y = \mu_{Y|X} + E = \alpha + \beta X + E \quad 2.3$$

Donde  $E$  es una variable aleatoria denominada término de error o perturbación estocástica, la cual se caracteriza por tener un valor esperado nulo, una varianza constante  $S^2$ , es independiente de la variable explicativa X y son independientes entre sí.

Si existen  $n$  pares de valores  $(x_i, y_i)$  de las variables aleatorias  $(X, Y)$  y están relacionadas linealmente por la función:

$$y_i = a + bx_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad 2.4$$

Entonces;

$$\varepsilon_i = y_i - \hat{y}_i = y_i - (a + bx_i) \quad 2.5$$





Donde

$\varepsilon_i$  Residuo y describe el error en el ajuste del modelo en el punto  $i$  de los datos.

Con las prioridades:

$$E[\varepsilon_i] = 0, E[(\varepsilon_i - \mu_\varepsilon)^2] = E[\varepsilon_i^2] = S_\varepsilon^2 \quad y \quad E[\varepsilon_i \varepsilon_j] = 0 \quad 2.6$$

El método de mínimos cuadrados entre los estimadores  $\hat{a}$  y  $\hat{b}$  de los parámetros de la expresión al minimizar la suma de los cuadrados de los errores y  $\varepsilon_i$ .

$$\hat{a}\hat{b}\varepsilon_i \min SS = \min \sum_{i=1}^n \varepsilon_i^2 = \min \sum_{i=1}^n [y_i - (a + bx_i)]^2 \quad 2.7$$

Para encontrar el mínimo se deriva la expresión con respecto a cada uno de los parámetros y se igualan con cero  $\delta SS/\delta b=0$ , así,

$$-2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \quad 2.8$$

$$-2 \sum_{i=1}^n (y_i - a - bx_i)x_i = 0 \quad 2.9$$

Que al resolver para el conjunto de parámetros se tiene

$$a = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \bar{y} - b\bar{x} \quad 2.10$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n \bar{y}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad 2.11$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad 2.12$$





$$S_x = \left[ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2} \quad S_x = \left[ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}$$

2.13

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad 2.14$$

$$S_y = \left[ \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2} \quad 2.15$$

La suma de los cuadrados y productos cruzados aparecen frecuentemente en cálculos de regresión, y se define como

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \left\{ \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 \right\} \quad 2.16$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \left\{ \sum_{i=1}^n y_i^2 - \frac{1}{n} (\sum_{i=1}^n y_i)^2 \right\} \quad 2.17$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \left\{ \sum_{i=1}^n x_i y_i - \frac{1}{n} (\sum_{i=1}^n x_i \sum_{i=1}^n y_i) \right\} \quad 2.18$$

Así,

$$b = S_{xy} / S_{xx} \quad 2.19$$

El coeficiente de correlación lineal simple se define por

$$r = \frac{\text{Cov}(x,y)}{\sqrt{S_x^2} \sqrt{S_y^2}} = \frac{S_{xy}[n-1]^{-1}}{S_x S_y}, \quad -1 < r < 1 \quad 2.20$$

El coeficiente de determinación, que es igual a la porción de la varianza de y que es explicada por la ecuación de regresión, se expresa como:





$$r^2 = b^2 \frac{S_x^2}{S_y^2}, \quad 0 < r^2 < 1 \quad 2.21$$

De esta última relación se desprende que para  $r^2 = 0$  la regresión no explica nada de la varianza de las variables independientes. Cuando  $r^2 = 1$  no existen desviaciones o errores en torno a la línea de mínimos cuadrados, es decir, se tiene un ajuste perfecto.

### 2.2.2 Regresión Lineal Múltiple

Un modelo de regresión lineal múltiple (MRLM) trata de explicar el comportamiento de una determinada variable dependiente o exógena ( $Y$ ) en función de un conjunto de  $k$  variables independientes  $X_1, X_2, \dots, X_k$  mediante una relación de dependencia lineal (suponiendo  $X_1 = 1$ ):

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U \quad 2.22$$

Donde

$U$  Término de error

Para determinar el modelo anterior, es necesario estimar el valor de los coeficientes  $\beta_1, \beta_2, \dots, \beta_k$ . La linealidad en parámetros posibilita la interpretación correcta de los parámetros del modelo. Los parámetros miden la intensidad media de los efectos de las variables independientes sobre la variable a explicar y se obtienen al tomar las derivadas parciales de la variable a explicar respecto a cada una de las variables independientes:

$$\beta_j = \frac{\partial Y}{\partial X_j}; j = 1, \dots, k \quad 2.23$$

El objetivo es asignar valores numéricos a los parámetros  $\beta_1, \beta_2, \dots, \beta_k$ . Es decir, se trata de estimar el modelo de manera que, los valores ajustados de la variable dependiente resulten tan próximos a los valores realmente observados como sea posible.





A fin de poder determinar las propiedades de los estimadores obtenidos al aplicar distintos métodos de estimación y realizar diferentes contrastes, se especifica un conjunto de hipótesis sobre el MRLM. Existen tres grupos de hipótesis:

- ✓ Hipótesis sobre el término del error
- ✓ Hipótesis sobre las variables independientes
- ✓ Hipótesis sobre los parámetros del modelo

### 2.2.2.1 Hipótesis sobre el término del error

Para una muestra de  $n$  observaciones se tendrá el siguiente sistema de  $n$  ecuaciones lineales:

$$\begin{aligned} Y_1 &= \beta_1 + \beta_2 X_{21} + \dots + \beta_k X_{k1} + U_1 \\ Y_2 &= \beta_1 + \beta_2 X_{22} + \dots + \beta_k X_{k2} + U_2 \\ &\vdots \\ Y_n &= \beta_1 + \beta_2 X_{2n} + \dots + \beta_k X_{kn} + U_n \end{aligned} \quad 2.24$$

o, en forma matricial:

$$Y = XB + U \quad 2.25$$

Donde:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \dots & \vdots \\ 1 & X_{2n} & \dots & X_{kn} \end{bmatrix}, \quad B = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$$

La hipótesis del MRLM se resume en el término de error, así:

a) El valor esperado del error es cero:

$$E[U_i] = 0 \quad ; \quad i = 1, \dots, n \quad 2.26$$





**b)** Todos los términos de error tienen la misma varianza (varianza constante):

$$\text{Var}[U_i] = \text{Var}[U_j] = \sigma^2 \quad ; \quad i \neq j \quad 2.27$$

Por lo tanto, todos los términos de la diagonal principal de la matriz de varianzas y covarianzas serán iguales:

$$\text{Var}[U] = \begin{bmatrix} \sigma^2 & \dots & \dots & \dots \\ \dots & \sigma^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \sigma^2 \end{bmatrix} \quad 2.28$$

**c)** No autocorrelación: los errores son independientes unos de otros, la matriz de varianzas y covarianzas es una matriz diagonal (fuera de la diagonal principal los elementos son cero):

$$\text{Var}[U] = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \quad 2.29$$

Bajo las hipótesis a) y c), la matriz de varianzas y covarianzas tendrá la siguiente forma:

$$\text{Var}[U] = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2(I_n) \quad 2.30$$

Donde

$I_n$  Matriz identidad de orden  $n$





d) El error sigue una distribución normal:

$$U \approx N(0_n, \sigma^2(I_n)) \quad 2.31$$

### 2.2.2.2 Hipótesis sobre las variables independientes

- Las variables independientes son fijas o determinísticas.
- La variable independientes no están correlacionadas con el error aleatorio.
- Las variables independientes no presentan relación lineal exacta entre sí.
- Las variables independientes son medidas sin error.
- En el modelo no se excluyen las variables relevantes e irrelevantes al explicar el comportamiento de la variable dependiente.

### 2.2.2.3 Hipótesis sobre los parámetros del modelo

La hipótesis de permanencia estructural, lo cual quiere decir que los parámetros poblacionales,  $\beta_1, \beta_2, \dots, \beta_k$ , a partir de la información muestral disponible de las variables observables del modelo. Únicamente se considerara dos métodos de estimación:

- ✓ Método de mínimos cuadrados ordinarios (MCO)
- ✓ Método de máxima verosimilitud (MV)

### Estimación por mínimos cuadrados (MCO)

Sea un modelo en forma matricial

$$Y = X(B) + U \quad 2.32$$

Suponiendo que el modelo ha sido estimado, obteniéndose  $\hat{Y}$ , vector de valores de la variable dependiente implicado por el modelo. La diferencia entre los valores observados y los estimados,  $e = Y - \hat{Y} = Y - X(\hat{B})$  se denomina vector de residuos. Ahora bien, el problema consiste en minimizar la suma de los cuadrados de residuos, con respecto del vector de parámetros





estimados,  $B$ . Del problema de optimización se deduce la siguiente expresión de mínimos cuadrados ordinarios del modelo de regresión lineal múltiple:

$$\hat{B} = (X'(X))^{-1}(X')Y \quad 2.33$$

Cuya varianza viene dada por

$$\text{var}[\hat{B}] = \sigma^2(X'(X))^{-1} \quad 2.34$$

El estimador MCO de la varianza del término de error es:

$$\hat{\sigma}_U^2 = \frac{e'e}{n-k} \quad 2.35$$

Donde  $n$  es el número de observaciones y  $K$  es el número de elementos del vector  $B$ .

Bajo la hipótesis de errores cíclicos, el estimador MCO del vector  $B$  cumple una serie de propiedades que le convierte en un insesgado (el valor esperado del estimador coincide con el valor real del parámetro), eficiente (de varianza mínima), y consistente.

Además, el estimados MCO de la varianza del término de error  $\hat{\sigma}_U^2$  es también insesgado.

### Estimación por máxima verosimilitud

Este método propone como un estimador el valor que maximiza la probabilidad de obtener la muestra ya disponible.

Se basa prácticamente en la distribución que sigue el término de error. A tales efectos, se suele suponer que los errores aleatorios tienen una distribución normal que, además de cumplir las propiedades de una muestra grande, es una aproximación cómoda y fácil de manejar.

El modelo que se utiliza es  $Y = X(B) + U$ , se supondrá que el término aleatorio sigue la distribución Normal:





$$f(u_i) = \frac{1}{2\pi\sigma} \exp\left\{-\frac{u_i^2}{2\sigma^2}\right\}, \quad i = 1, \dots, n \quad 2.36$$

Maximizar la posibilidad de obtener la muestra ya disponible equivale a maximizar la función de densidad conjunta del vector aleatorio,  $U$ . Por lo tanto la expresión de la función de densidad conjunta es:

$$f(U) = \prod_{i=1}^n f(u_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{\sum u_i^2}{2\sigma^2}\right\} \quad 2.37$$

Como  $U$  sigue una distribución Normal Multivariada de orden  $k$ , la variable  $Y$ , al ser una combinación lineal de los errores aleatorios, también seguirá el comportamiento de la distribución Normal Multivariada. Así para que la función de densidad conjunta sea una función de verosimilitud, el vector aleatorio  $U$  se expresara en función del vector  $Y$ , es decir:

$$L(Y; \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right\} \quad 2.38$$

Se trata de maximizar la función de verosimilitud. Como la ecuación anterior resulta complicada, se aplicará una transformación logarítmica:

$$\ln L(Y; \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(2\sigma^2) - \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \quad 2.39$$

Derivando la función de verosimilitud con respecto a  $B$  y  $\sigma^2$ , e igualando las derivadas a cero, se obtiene:

$$\hat{B}_{MV} = (X'X)^{-1} X'Y \quad 2.40$$

Cuya varianza es la siguiente:

$$\text{var}[\hat{B}_{MV}] = \sigma^2 (X'X)^{-1} \quad 2.41$$





Además el estimador MCO de la varianza del término de error es:

$$\hat{\sigma}_{MV}^2 = \frac{e'e}{n-k} \quad 2.42$$

Donde  $n$  es el número de observaciones y  $k$  es el número de elementos del vector  $B$ .

Se observa que el estimador de MV de  $B$  coincide con el MCO, con lo que se tendrá las mismas propiedades; será lineal, insesgado, óptimo y consistente.

Es fácil ver que el estimador de MV de  $\sigma^2$ , en cambio, resulta diferente del MCO y no es insesgado aunque si es asintóticamente insesgado.

#### 2.2.2.4 Medidas de Bondad de Ajuste

Una de las técnicas es la suma de los cuadrados de errores, SCE, que puede expresarse de varias formas, una de ellas es:

$$e'e = \sum_{i=1}^n e_i^2 = Y'Y - \hat{B}X'Y = Y'Y - \hat{Y}'\hat{Y} = \sum_{i=1}^n Y_i^2 - \sum_{i=1}^n \hat{Y}_i^2 \quad 2.43$$

Despejando la suma de cuadrados de la variable dependiente se tiene:

$$Y'Y = \hat{Y}'\hat{Y} + e'e \quad 2.44$$

O bien

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n \hat{Y}_i^2 + \sum_{i=1}^n e_i^2 \quad 2.45$$

Restando a ambos lados el término  $n\bar{Y}^2$  se obtiene:

$$Y'Y - n\bar{Y}^2 = \hat{Y}'\hat{Y} - n\bar{Y}^2 + e'e \quad 2.46$$

O bien:





$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n e_i^2 \quad 2.47$$

El miembro izquierdo representa la suma de cuadrados totales (SCT) y no es sino la suma de cuadrados de las desviaciones respecto a su media aritmética.

Por otra parte, si el modelo tiene término independiente, a la cantidad  $\hat{Y}'\hat{Y} - n\bar{Y}^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$  se le denomina suma de cuadrados de la regresión (SCR).

Se definirá el coeficiente de determinación,  $R^2$ , el cual será la primera medida de bondad de ajuste:

$$R^2 = 1 - \frac{SCE}{SCT} \quad 2.48$$

Donde:

SCE Suma de los errores cuadrados

SCT Variación Total de la variable dependiente

Si el modelo tiene término independiente, entonces se cumple la igualdad  $SCT=SCR+SCE$ , y el coeficiente de determinación podrá expresarse como:

$$R^2 = \frac{SCR}{SCT} \quad 2.49$$

El coeficiente de determinación indica que proporción de variabilidad total queda explicada por la regresión. Si el modelo tiene término independiente, entonces  $R^2$  toma valores entre 0 y 1.

El uso de  $R^2$  presenta algunas limitaciones a la hora de comparar varios modelos desde la perspectiva de bondad de ajuste. Cuando se desea llevar a cabo un análisis comparativo entre varios modelos, se utilizará el valor de  $R^2$  corregido.





$$R^2 = 1 - \frac{n-1}{n-k} (1 - R^2)$$

2.50

### 2.2.2.5 Análisis estadístico

Para contrastar las hipótesis de significatividad individual, se tiene:

$$H_0 : \beta_j = 0 \quad 2.51$$

$$H_A : \beta_j \neq 0 \quad 2.52$$

El estadístico t-Student que se utiliza para realizar la prueba es:

$$t_j = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}_u^2 \alpha_{jj}}} \cong t_{n-k} \quad 2.53$$

Donde

$$\sqrt{\hat{\sigma}_u^2 \alpha_{jj}} \quad \text{Error estimado de } \hat{\beta}_j$$

$$\alpha_{jj} \quad \text{J-ésimo elemento de la diagonal principal de la matriz } (X'X)^{-1}$$

Dado un nivel de significación de  $\alpha$ ; las tablas de distribuciones nos proporcionan la cantidad  $t_{n-k, \alpha/2}$  que es el valor asociado a una t-student con  $n-k$  grados de libertad que deja a su derecha un área de  $\alpha/2$  (o equivalente, deja a su izquierda un área de  $1-\alpha/2$ ). La regla de decisión que se utiliza para determinar si el parámetro asociado a la variable  $X_j$  es individualmente significativo o no será:

✓ Si  $|t_j| \geq t_{n-k, \alpha/2}$ , el estadístico cae fuera de la región de aceptación, por lo que se rechazará la hipótesis nula. Se concluirá que el parámetro es significativamente diferente de cero.





✓ Si  $|t_j| < t_{n-k, \alpha/2}$ , el estadístico cae dentro de la región de aceptación, por lo que no se rechazará la hipótesis nula. Por lo tanto, el parámetro no es significativo.

Si se desea contrastar la significación conjunta, las hipótesis se plantearan de la siguiente manera:

$$H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0 \quad 2.54$$

$$H_A : \text{NO } H_0 \quad 2.55$$

El término independiente no contribuye al explicar la variabilidad de la variable dependiente, con lo cual no se incluye en la restricción

El estadístico  $F$  de *Snedecor* que se utilizará para la realización de la prueba será:

$$F_0 = \frac{R^2}{1-R^2} \left( \frac{n-k}{n-1} \right) \cong F_{k-1, n-k} \quad 2.56$$

El estadístico se distribuye bajo la hipótesis nula con una distribución  $F$  de *Snedecor* con  $k-1$  grados de libertad en el numerador y  $n-k$  grados de libertad en el denominador. La regla de decisión utilizada para comprobar la significación global del modelo será:

✓ Si  $F_0 \geq F_{k-1, N-k; \beta\alpha}$  el estadístico de contraste cae fuera de la región de aceptación, con lo que se rechaza la hipótesis nula. Por lo tanto es globalmente significativo.

✓ Si  $F_0 < F_{k-1, N-k; \beta\alpha}$ , el estadístico de contraste cae dentro de la región de aceptación, de modo que ahora la hipótesis nula no se rechaza. En consecuencia, se puede afirmar que el modelo no es globalmente significativo.





## 2.3 Funciones de probabilidad

### 2.3.1 Distribución exponencial con un parámetro $\beta$

$$f(x) = 1 - \beta e^{-\beta x} \quad x > 0 \quad 2.57$$

Donde

$\beta$  Parámetro de escala

Además

$$\mu = \frac{1}{\beta} \quad 2.58$$

$$\sigma^2 = \frac{1}{\beta^2} \quad 2.59$$

### 2.3.2 Distribución exponencial de dos parámetros $\beta$ y $x_0$

$$f(x) = 1 - e^{-\left(\frac{x-x_0}{\beta}\right)} \quad 2.60$$

$$f(x) = \frac{1}{\beta} e^{-\left(\frac{x-x_0}{\beta}\right)} \quad 2.61$$

Donde

$x_0$  Parámetro de ubicación

$\beta$  Parámetro de escala

Además

$$\mu = x_0 + \beta \quad 2.62$$

$$\sigma^2 = \beta^2 \quad 2.63$$

$$\gamma = 2 \quad 2.64$$





### 2.3.3 Distribución Normal

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad 2.65$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad 2.66$$

Donde

$\mu$  *parametro de ubicación* Parámetro de ubicación  
 $\sigma$  Parámetro de escala

Además

$$E(x) = \mu \quad 2.67$$

$$E[(x - \mu)^2](x) = \sigma^2 \quad 2.68$$

$$\gamma = 0 \quad 2.69$$

$$k = 3 \quad 2.70$$

$$M(t) = e^{\mu t + (\sigma^2 t^2 / 2)} \quad 2.71$$

$$C(\theta) = e^{i\mu\theta - (\sigma^2 \theta^2 / 2)} \quad 2.72$$

### 2.3.4 Distribución Log Normal con dos parámetros

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln(x)-\mu_y}{\sigma_y}\right]^2}, x > 0 \quad 2.73$$

Donde

$\mu_y$  Parámetro de ubicación

$\sigma_y$  Parámetro de escala

$\gamma > 0$

### 2.3.5 Distribución Log Normal con tres parámetros





$$f(x) = \frac{1}{(x-x_0)} e^{-\frac{1}{2} \left[ \frac{\ln(x-x_0) - \mu_y}{\sigma_y} \right]^2}, x > x_0 \quad 2.74$$

Donde

$X_0$  Parámetro de ubicación

$\mu_y$  Parámetro de forma

$\sigma_y$  Parámetro de escala

### 2.3.6 Distribución Gamma con dos parámetros

$$F(x) = \int_0^x \frac{x^{\beta-1} e^{-x/\alpha}}{\alpha^{\beta} \Gamma(\beta)} dx \quad 2.75$$

¡Error! Se espera un dígito.

si  $\alpha > 0 \rightarrow \gamma > 0$

Donde

$\alpha$  Parámetro de escala

$\beta$  Parámetro de forma

$\Gamma(b)$  Función Gamma completa

Además

$$\mu = \beta\alpha \quad 2.77$$

$$\sigma^2 = \alpha^2\beta \quad 2.78$$

$$\gamma = \frac{2}{\sqrt{\beta}} \quad 2.79$$

### 2.3.7 Distribución Gamma con tres parámetros

$$f(x) = \frac{1}{\alpha^{\beta} \Gamma(\beta)} \left( \frac{x-x_0}{\alpha} \right)^{\beta-1} e^{-\left( \frac{x-x_0}{\alpha} \right)}, \alpha > 0 \quad 2.80$$





$$x_0 \leq x < \infty$$

¡Error! Se espera un dígito.

Donde

$X_0$  Parámetro de ubicación

$\alpha$  Parámetro de escala

$\beta$  Parámetro de forma

Además

$$\mu = x_0 + \alpha\beta \quad 2.81$$

$$\sigma^2 = \alpha^2\beta \quad 2.82$$

$$\gamma = \frac{2}{\sqrt{\beta}} \quad 2.83$$

$$k = 3\left(1 + \frac{\gamma^2}{2}\right) \quad 2.84$$

### 2.3.8 Distribución de valores extremos tipo I (Gumbel)

$$F(x) = e^{-e^{-\left[\frac{x-v}{\alpha}\right]}} \quad 2.85$$

$$f(x) = \frac{1}{\alpha} e^{\left(\frac{x-v}{\alpha}\right)} e^{-e^{-\left(\frac{x-v}{\alpha}\right)}} \quad -\infty < x < \infty, \alpha > 0 \quad 2.86$$

Donde

$v$  Parámetro de ubicación

$\alpha$  Parámetro de escala

$$E(x) = \hat{v} + 0.5772\hat{\alpha} \quad 2.87$$

$$\sigma^2 = \frac{\pi^2 \hat{\alpha}^2}{6} \quad 2.88$$

$$\gamma = 1.1396 \quad 2.89$$





$$k = 5.4002 \quad 2.90$$

La variable reducida Gumbel es

$$y_i = \frac{x_i - v}{\alpha} \quad 2.91$$

## 2.4 Estimación de parámetros por el Método de Momentos

El método de los momentos es un procedimiento muy sencillo para encontrar un estimador de uno o más parámetros poblacionales. Consiste básicamente en plantear un sistema de ecuaciones, cuyo tamaño depende del número de parámetros a estimar. Esto se hace al igualar los momentos poblacionales con los muestrales.

Los momentos poblacionales pueden obtenerse con respecto a la media o con respecto al origen. Ya sea que se utilice una u otra se podrá hacer las transformaciones necesarias.

Los momentos muestrales, también conocidos como estadísticos muestrales, se obtienen con las siguientes expresiones.

### **Media**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad 2.92$$

### **Varianza Sesgada**

$$S_{sesg}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad 2.93$$

### **Varianza no sesgada**

$$S_{insesg}^2 = \frac{n}{n-1} S_{sesg}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad 2.94$$



**Coeficiente de asimetría sesgado**

$$g_{sesg} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{(S_{sesg}^2)^{\frac{3}{2}}} \quad 2.95$$

**Coeficiente de asimetría no sesgado**

$$g_{insesg} = \frac{n^2}{(n-1)(n-2)} g_{sesg} \quad 2.96$$

**Coeficiente de curtosis sesgado**

$$k_{sesg} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{(S_{sesg}^2)^2} \quad 2.97$$

**Coeficiente de curtosis no sesgado**

$$k_{insesg} = \frac{n^3}{(n-1)(n-2)(n-3)} k_{sesg} \quad 2.98$$

**Desviación estándar**

$$S = \sqrt{S^2} \quad 2.99$$

**Coeficiente de variación**

$$Cv = \frac{S}{\bar{x}} \quad 2.100$$

En el análisis hidrológico se recomienda el uso de los estadísticos no sesgados, ya que generalmente se trabaja con muestras relativamente pequeñas.





### 2.4.1 Distribución exponencial con un parámetro ( $\beta$ )

$$\beta = \frac{1}{\hat{\mu}} = \frac{1}{\bar{x}} \quad 2.101$$

### 2.4.2 Distribución exponencial de dos parámetros $\beta$ y $x_0$

$$\hat{\beta} = S \quad 2.102$$

$$\hat{x}_0 = \bar{x} - S \quad 2.103$$

### 2.4.3 Distribución Normal

$$\hat{\mu} = \bar{x} \quad 2.104$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad 2.105$$

### 2.4.4 Distribución Log Normal con dos parámetros

$$\hat{\mu}_y = \frac{1}{n} \sum_{i=1}^n \ln x_i \quad 2.106$$

$$\hat{\mu}_y^2 = \frac{\sum_{i=1}^n (\ln x_i - \hat{\mu}_y)^2}{n} \quad 2.107$$

### 2.4.5 Distribución Log Normal con tres parámetros

$$\hat{x}_0 = \bar{x} \left(1 - \frac{\hat{\eta}_x}{\hat{\eta}_s}\right) \quad 2.108$$

Donde

$$\hat{\eta}_x = \frac{S}{\bar{x}} \quad 2.109$$





$$\hat{\eta}_z = \frac{1 - w^{\frac{2}{3}}}{\frac{1}{w^{\frac{2}{3}}}} \quad 2.110$$

$$w = \frac{(g^2 + 4)^{\frac{1}{2}} - g}{2} \quad 2.111$$

$$\hat{\mu}_y = \ln\left(\frac{S}{\hat{\eta}_z}\right) - \frac{1}{2} \ln(\hat{\eta}_z^2 + 1) \quad 2.112$$

$$\hat{\sigma}_y = [\ln(\hat{\eta}_z^2 + 1)]^{\left(\frac{1}{2}\right)} \quad 2.113$$

#### 2.4.6 Distribución Gamma con dos parámetros

$$\hat{\alpha} = \frac{s^2}{\bar{x}} \quad 2.114$$

$$\hat{\beta} = \left(\frac{\bar{x}}{S}\right)^2 \quad 2.115$$

#### 2.4.7 Distribución Gamma con tres parámetros

$$\hat{\beta} = \frac{4}{g^2} \quad 2.116$$

$$\hat{\alpha} = \frac{S}{\sqrt{\hat{\beta}}} \quad 2.117$$

$$\hat{x}_0 = \bar{X} - S\sqrt{\hat{\beta}} \quad 2.118$$

#### 2.4.8 Distribución de valores extremos tipo I (Gumbel)

$$\hat{\theta} = \bar{x} - 0.45 S \quad 2.119$$





$$d = \frac{\sqrt{6}}{\pi} S = 0.785$$

2.120

## 2.5 Límites de confianza

Los límites de confianza son empleados para estimar las incertidumbres asociadas con la determinación de los eventos para periodos de retorno específicos.

Puesto que una distribución de frecuencias es únicamente un estimado de la muestra de cierta población, es probable que otra muestra de igual longitud de esa misma población, pero tomada en diferente tiempo produzca otra curva de frecuencias. Los límites o intervalos de confianza definen el rango dentro del cual se espera que se ubiquen éstas curvas con cierto nivel de confianza. Es decir,

$$\bar{X}_l = \bar{X}_T \pm u_\alpha S_T$$

4.1

2.121

Donde

$\bar{X}_l$  Límites de confianza superior e inferior

$U_\alpha$  Evento obtenido a partir de la función de distribución para cierto periodo de retorno

$U_\alpha$  Desviación normal estándar para un nivel de confianza

Con límites al 90%

$$\alpha = 0.10 \quad u_\alpha = 1.645212$$

Con límites al 95%

$$\alpha = 0.05 \quad u_\alpha = 1.960395$$

Con límites al 99%

$$\alpha = 0.01 \quad u_\alpha = 2.576236$$





$S_T$  Desviación estándar de los eventos estimados para un periodo de retorno  $T$ .

A continuación se dan las expresiones de  $\bar{X}_T$  y  $\bar{S}_T$  para algunas distribuciones (Kite, 1988).

### 2.5.1 Distribución Normal

$$\bar{X}_T = \bar{\mu} + \bar{\sigma}U_T \quad 2.122$$

Desviación estándar de los eventos  $\bar{X}_T$  por momentos y máxima verosimilitud.

$$S_T = \left[1 + \frac{U_T^2}{2}\right]^{1/2} \frac{\bar{\sigma}}{\sqrt{n}} \quad 4.3$$

2.123

Donde  $n$  es el tamaño de la muestra analizada.

Para un probabilidad acumulada  $0 < F(X) \leq 0.5$

$$U_T = \frac{b_0 + b_1v + b_2v^2}{1 + b_3v + b_4v^2 + b_5v^3} \quad 2.124$$

Donde

$$b_0 = 2.515517$$

$$b_1 = 0.802853$$

$$b_2 = 0.010328$$

$$b_3 = 1.432788$$

$$b_4 = 0.18926$$

$$b_5 = 0.001308$$

$$v = \sqrt{\ln\left\{\frac{1}{[F(x)]^2}\right\}} \quad 2.125$$





Para una probabilidad acumulada  $0.5 < F(x) \leq 1 < F(x) \leq 1$  se cambia  $F(x)$  por  $[1 - F(x)]$   $[1 - F(x)]$  en la expresión 2.125 y el signo al valor  $U_T$  calculado con la ecuación 2.124 aquí  $F(x) = 1/T$  y  $T =$  periodo de retorno en años.

### 2.5.2 Distribución Log Normal con 2 parámetros

$$\hat{X}_T = \exp(\hat{\mu}_y + U_T \hat{\sigma}) \quad 2.126$$

Desviación estándar los de los eventos  $\hat{X}_T$  por momentos

$$S_T = \left(1 + \frac{U_T^2}{2}\right)^{1/2} \frac{\sigma_y}{\sqrt{n}} \quad 2.127$$

### 2.5.3 Distribución de valores extremos tipo I (Gumbel)

$$\hat{X}_T = \hat{\nu} - \hat{\theta} \ln[-\ln(1 - 1/T)] \quad 4.38$$

Desviación estándar de los eventos  $\hat{X}_T$  por momentos

$$S_T = \left[ \frac{\hat{\theta}^2}{2} (1 + 1.1396 k_T + 1.10 k_T^2) \right]^{1/2} \quad 2.129$$

Donde

$$k_T = - \left\{ 0.45 + 0.7797 \ln \left[ -\ln \left( 1 - \frac{1}{T} \right) \right] \right\} \text{ donde } T = \frac{n+1}{m} \quad 2.130$$

## 2.6 Cálculo de homogeneidad de una muestra

Las fases de planeación, diseño, construcción y operación de los aprovechamientos hidráulicos están siempre relacionadas con eventos hidrológicos futuros. La complejidad de los procesos físicos de estos eventos





hace casi imposible tener estimaciones confiables de diseño basadas en las leyes de la mecánica o la física, ya sea porque estos métodos son insuficientes o porque el modelo matemático resultante es muy complicado. Una alternativa en el análisis hidrológico es la aplicación de los conceptos de la teoría de probabilidad y estadística.

El análisis de frecuencias de los gastos máximos anuales se emplea para proveer la magnitud de un evento  $Q_T$ , de cierto periodo de retorno  $T$ , para el diseño de una obra hidráulica; manejo de las llanuras de inundación, y como ayuda en los procesos de planeación y manejo de las cuencas hidrológicas. Sin embargo, el proyectista no solo debe estimar la magnitud del evento del diseño, sino también debe proporcionar la excedencia, con el fin de fijar la seguridad del funcionamiento de la obra, o bien el riesgo de falla.

Las características estadísticas de las series hidrológicas, como la media, la desviación estándar y los coeficientes de correlación serial, se afectan cuando la serie presenta tendencia en la media o en la varianza, o cuando ocurren saltos negativos o positivos; tales anomalías son producidas por la pérdida de homogeneidad.

En general, la falta de homogeneidad de los datos es inducida por las actividades humanas como la deforestación, apertura de nuevas áreas de cultivo, rectificación de cauces, construcción de embalses y reforestación. También es producto de los procesos naturales súbitos, como incendios forestales, terremotos, deslizamientos de laderas y erupciones volcánicas.

Las pruebas estadísticas que miden la homogeneidad de una serie de datos presentan una hipótesis nula y una regla para aceptarla o rechazarla.

A continuación se describen tres tipos de pruebas para la determinación de la homogeneidad de una muestra.

### 2.6.1 Prueba estadística de Helmert

Esta prueba es sencilla y consiste en analizar el signo de las desviaciones de cada evento  $Q_j$  de la serie para  $j$  para  $i=1,2,\dots,n_j$ , con respecto al valor medio  $Q_j$ . Si una desviación de un cierto signo es seguida de otra del mismo





signo, entonces se dice que se forma una secuencia  $S$ , de lo contrario se considera como un cambio  $C$ .

La serie se considera homogénea si se cumple :

$$-\sqrt{n_j - 1} \leq (S - C) \leq \sqrt{n_j - 1} \quad 2.131$$

### 2.6.2 Prueba estadística $t$ de Student

Cuando la causa probable de la perdida de la homogeneidad de la serie sea un camio abrupto en la media, la prueba del estadístico  $t$  es muy útil.

Si se considera una serie  $Q_j$  para  $i=1,2,\dots,n$ , del sitio  $j$ , la cual se divide en dos conjuntos de tamaño  $n_1=n_2=n_j/2$ , entonces, el estadístico de prueba se define con la expresión:

$$t_d = \frac{\bar{x}_1 - \bar{x}_2}{\left[ \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{1/2}}$$

Donde

- ✓  $\bar{x}_1, S_1^2$  son la media y la varianza de la primera parte del registro de tamaño  $n_1$ .
- ✓  $\bar{x}_2, S_2^2$  son la media y la varianza de la primera parte del registro de tamaño  $n_2$ .

El valor absoluto de  $t_d$  se compra con el valor de la distribución  $t$  de Student de dos colas, y con  $\nu = n_1 + n_2 - 2$  grados de libertad y para un nivel  $\alpha = 0.05$ , tabla 2.1.





## DISTRIBUCIÓN *t* DE STUDENT

Grados de libertad	Nivel de significancia	
	Una cola 5%	Dos colas 5%
1	6.314	12.706
2	2.920	4.303
3	2.353	3.182
4	2.132	2.776
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.812	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145
15	1.753	2.131
16	1.746	2.120
17	1.740	2.110

Grados de libertad	Nivel de significancia	
	Una cola 5%	Dos colas 5%
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
26	1.706	2.056
27	1.703	2.052
28	1.701	2.048
29	1.699	2.045
30	1.697	2.042
31	1.684	2.021
32	1.671	2.000
33	1.658	1.980
34	1.645	1.960

**Tabla 2.1** Distribución *t* de Student para una y dos colas y  $\alpha = 5\%$

Si y solo sí, el valor absoluto de  $t_d$  es mayor que aquel de la distribución *t* de Student, se concluye que la diferencia entre las medias es evidencia de inconsistencia y por lo tanto la serie  $Q_j$  se considera homogénea.





### 2.6.3 Prueba estadística de Cramer

Esta prueba se utiliza con el propósito de verificar homogeneidad en el registro  $Q_i^j$  de la serie  $j$  para  $i=1,2,\dots,n_j$ , y también para determinar si el valor medio no varía significativamente de un periodo de tiempo a otro. Con este propósito se consideran tres bloques, el primero, del tamaño total de la muestra  $n_j$ ; el segundo de tamaño  $n_{60}$  (60% de los últimos valores de la muestra  $n_j$ ); y el tercero de tamaño  $n_{30}$  (30% de los últimos valores de la muestra  $n_j$ ).

La prueba compara que el valor de la media de  $Q_i^j$  del registro total con cada una de las medidas de los bloques elegidos  $Q_{60}^j$  y  $Q_{30}^j$ . Para que se considere la serie analizada como estacionaria en la media, se deberá cumplir que no existe una diferencia significativa entre las medias de los bloques.

$$\bar{Q}^j = \sum_{i=1}^{n_j} \frac{Q_i^j}{n_j}, \quad 2.133$$

$$s_{Q^j} = \left[ \frac{1}{(n_j)} \sum_{i=1}^{n_j} (Q_i^j - \bar{Q}^j)^2 \right]^{\frac{1}{2}} \quad 2.134$$

para una sola muestra analizada  $j = 1$

$$\bar{Q}_{60}^j = \sum_{k=1}^{n_{60}} \frac{Q_k^j}{n_{60}} \quad 2.135$$

$$\bar{Q}_{30}^j = \sum_{k=1}^{n_{30}} \frac{Q_k^j}{n_{30}} \quad 2.136$$

$$T_{60}^j = \frac{\bar{Q}_{60}^j - \bar{Q}^j}{s_{Q^j}} \quad 2.137$$





$$T_{30}^j = \frac{Q_{30}^{j} - \bar{Q}^j}{S_{Q^j}}$$

$$t_w = \left\{ \frac{n_w(n_j - 2)}{n_j - n_w[1 + (\bar{x}_w^j)^2]} \right\}^{\frac{1}{2}} |\bar{x}_w^j| \quad \text{para } w=60 \text{ y } w=30 \quad 2.139$$

El estadístico  $t_w$  tiene distribución  $t$  de Student de dos colas con  $v = n_1 + n_2 - 2$  grados de libertad y para un nivel  $\alpha = 0.05$ .

Si y solo si, el valor absoluto de  $t_w$ , para  $w=60$  y  $w=30$ , es mayor que el de la distribución  $t$  de Student, se concluye que la diferencia entre las medias es evidencia de inconsistencia y por tanto la serie  $Q^j$  se considera no homogénea.

## 2.7 Determinación de independencia de una muestra

Para que se pueda llevar a cabo el análisis de frecuencias se requiere que la muestra  $Q^j$  de la serie  $j$  para  $i = 1, 2, \dots, n_j$ , este compuesta por variables aleatorias. Para probarlo se aplica la prueba de independencia de Anderson, la cual hace uso del coeficiente de autocorrelación serial  $r_k^j$  para diferentes tiempos de retraso  $k$ . Si se analiza un solo registro, entonces  $j=1$ .

La expresión para obtener el coeficiente de autocorrelación serial de retraso  $k$  es:

$$r_k^j = \frac{\sum_{i=1}^{n_j} (q_i^j - \bar{q}^j)(q_{i+k}^j - \bar{q}^j)}{\sum_{i=1}^{n_j} (q_i^j - \bar{q}^j)^2}; \quad r_0^j = 1 \text{ y } k = 1, 2, \dots, \frac{n_j}{3} \quad 2.140$$

Donde

$$\bar{q}^j = \sum_{i=1}^{n_j} \frac{q_i^j}{n_j} \quad 2.141$$

Además los límites al 95% de confianza para  $r_k^j$  se pueden obtener como





$$r_{k}^{j}(95\%) = \frac{-1 \pm 1.96 \sqrt{(n_j - k - 1)}}{n_j - k}$$

2.142

La gráfica de los valores estimados para  $r_{k}^{j}$  (ordenadas) contra los tiempos de retraso  $k$  (abscisas), junto con sus correspondientes límites de confianza, se llama correlograma de la muestra.

Si y solo si, el 10% de los valores de  $r_{k}^{j}$  sobrepasan los límites de confianza se dice que la serie  $Q_i^j$  es independiente y por lo tanto es una variable que sigue las leyes de probabilidad.

## 2.8 Prueba de bondad de ajuste

Kite (1988) propuso un estadístico que permite seleccionar la mejor opción, entre diferentes modelos en competencia, para el ajuste de una muestra de datos  $Q_i^j, Q_i^j$  para  $i=1,2,\dots,n_j$ , de un sitio  $j$ .

Este estadístico es conocido como el error estándar de ajuste, y tiene la forma

$$EE = \left[ \sum_{i=1}^{n_j} \frac{(\hat{Q}_T^j - Q_T^j)^2}{n_j - mp} \right]^{1/2} \quad 2.143$$

Donde  $Q_T^j$  son los eventos  $Q_i^j$  ordenadas de mayor a menor con un periodo de retorno asignado con la ecuación de Weibull como:

$$T = \frac{n_j + 1}{m} \quad 2.144$$

y una probabilidad de no excedencia:





$$P = 1 - \frac{1}{T} \quad 2.145$$

$n_j$  Longitud en años del registro analizado.

$m$  Número de orden del registro,  $m=1$  para el evento más grande y  $m=n_j$  para evento más chico.

$Q_T^j$  Eventos estimados por cierta distribución de probabilidad para cada periodo de retorno  $T$  asignado a la muestra ordenada  $Q_i^j$ .

$mp$  Número de parámetros de la distribución ajustada, donde

La distribución de mejor ajuste será aquella que proporcione el mínimo valor del estadístico  $EE$ . Si una o mas distribuciones tienen valores similares del  $EE$ , entonces, se deberá optar por aquella distribución que tenga el menor número de parámetros.

## 2.9 Análisis de frecuencias de gastos máximos anuales

El análisis de frecuencias de los gastos máximos anuales de una muestra  $Q_i$ ,  $i=1,2,\dots,n$ , se emplea para proveer la magnitud de un evento  $\hat{Q}_T$ , de cierto periodo de retorno  $T$ , por medio del ajuste de una distribución de probabilidad, la cual es seleccionada como la mejor de un grupo de ellas.

La secuencia de análisis es la siguiente:

**Paso 1.** Recabar la información de los eventos  $Q_i$ ,  $i=1,2,\dots,n$ . En este punto se debe verificar la calidad y cantidad de la información.

**Paso 2.** Verificar la homogeneidad de la serie mediante las pruebas de Helmer,  $t$  de Student y Cramer.

**Paso 3.** Obtención de los estadísticos muestrales de la serie  $Q_i$  prefiriendo los no sesgados, dado que generalmente se trabaja con muestras pequeñas.

**Paso 4.** Verificar con la prueba de Anderson la independendencia de eventos de la serie  $Q_i$ .





**Paso 5.** La serie  $Q_i$  se ordena de mayor a menor, se le asigna un periodo de retorno  $T$  y una probabilidad de no excedencia.

**Paso 6.** A la serie  $Q_i$  se le ajustan las diferentes distribuciones de probabilidad para el análisis de máximos y se selecciona aquel que proporcione el mínimo error estándar de ajuste  $EE$ .

**Paso 7.** Una vez que se obtiene la distribución de mejor ajuste de registro  $Q_i$ , es posible calcular los eventos  $\hat{Q}_T$  y sus límites de confianza para los periodos de retorno  $T= 2, 5, 10, 20, 50, 100, 500, 1000, 5000$  y  $10,000$  años.

