

HIDROLOGÍA ESTADÍSTICA



Edición 2025

Alejandra De Vera (en base a notas de Rafael Terra)

Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA)
Facultad de Ingeniería, Universidad de la República, Uruguay

adevera@fing.edu.uy

DISEÑO HIDROLÓGICO

Concepto

- ❖ Proceso de **evaluación del impacto** de los eventos hidrológicos en un sistema hídrico y de **selección de las magnitudes de diseño** de las variables más importantes del sistema, para que este se comporte adecuadamente.
- ❖ Se basa en **describir cabalmente el comportamiento del sistema**, incluidas todas sus **propiedades estadísticas** (media, dispersión, eventos extremos de máximas y de mínimas, etc.), así como la **variabilidad** innata de los procesos hidrológicos/meteorológicos involucrados.



HIDROLOGÍA ESTADÍSTICA

Concepto

Interpretación de registros hidrológicos históricos, rara vez completos y que contienen errores, de modo de **caracterizar el comportamiento de los estadísticos más relevantes** de las variables hidrológicas con el objetivo del diseño hidrológico.

El objetivo de la estadística es **extraer información** esencial de un **conjunto de datos**, reduciendo un conjunto grande de números a un conjunto pequeño de números.

Hidrología Aplicada

Ven Te Chow (1994), David R. Maidment y Larry W. Mays

Capítulo 11: Hidrología Estadística

HIDROLOGÍA ESTADÍSTICA

Concepto

Interpretación de registros hidrológicos históricos, rara vez completos y que contienen errores, de modo de **caracterizar el comportamiento de los estadísticos más relevantes** de las variables hidrológicas con el objetivo del diseño hidrológico.

El **objetivo del diseño hidrológico es clave** para definir qué características y propiedades son de interés y por lo tanto qué “conjunto” de herramientas estadísticas es pertinente aplicar.

Hidrología Aplicada

Ven Te Chow (1994), David R. Maidment y Larry W. Mays

Capítulo 11: Hidrología Estadística

HIDROLOGÍA ESTADÍSTICA

Diseño Hidrológico

- ❖ Puentes y alcantarillas
- ❖ Control de avenidas e inundaciones
- ❖ Estructuras para almacenar agua
- ❖ Abastecimiento de agua a ciudades e industrias
- ❖ Sistemas y obras de riego
- ❖ Sistemas de drenaje
- ❖ Recolección, tratamiento y disposición de aguas residuales
- ❖ Gestión integral de los recursos hídricos



HIDROLOGÍA ESTADÍSTICA

Índice

- ❖ **Caracterización de la información hidrológica**
 - ✓ Cuantiles de una muestra
 - ✓ Principales medidas numéricas de resumen
 - ✓ Representaciones gráficas
- ❖ **Regresión entre dos variables**
 - ✓ Una variable hidrológica y el tiempo
 - ✓ Dos variables apareadas
 - ✓ Regresión lineal simple
 - ✓ Coeficiente de correlación
- ❖ **Control de calidad de datos (¿Contexto?)**
- ❖ **Curva de permanencia de caudales**
- ❖ **Análisis de extremos (Clases Rafael)**

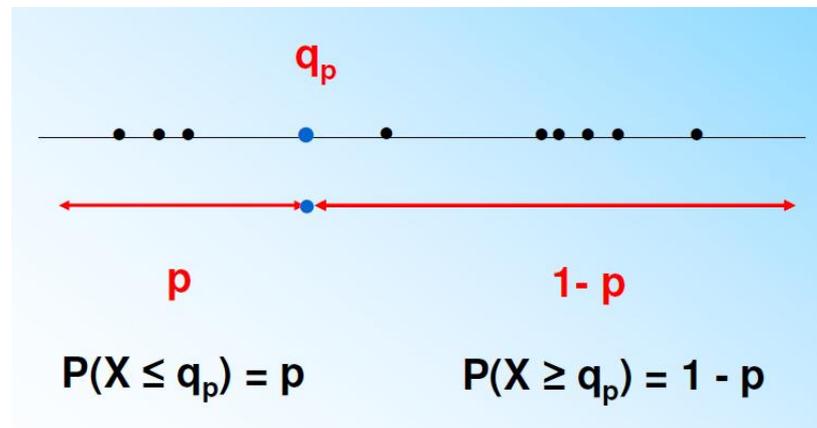
CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Cuantiles de una muestra aleatoria

(Percentiles, terciles, cuartiles, quintiles, deciles, etc.)

Los **p-quantiles** (q_p) (o percentiles) son valores que dejan, en cierto sentido, probabilidad p a su izquierda, y probabilidad $1-p$ a su derecha (sea p tal que $0 < p < 1$).

En general, los percentiles no son únicos y, por lo tanto, no hay una única forma de estimarlos.



CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Cuantiles de una muestra aleatoria

(Percentiles, terciles, cuartiles, quintiles, deciles, etc.)

Una forma posible para una muestra aleatoria de tamaño n es:

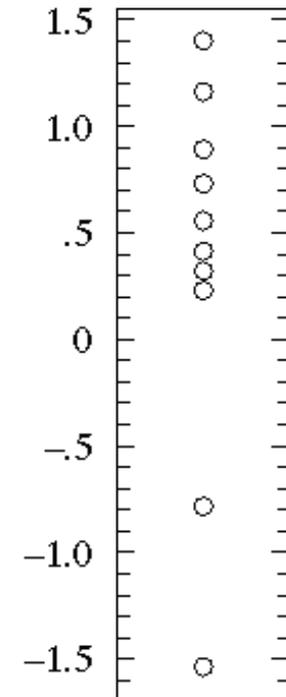
- 1) Tomar los estadísticos de orden como los cuantiles $(0.5/n)$, $(1.5/n)$, ..., $([n-0.5]/n)$ respectivamente
- 2) Para los cuantiles con probabilidades entre $(0.5/n)$ y $([n-0.5]/n)$, se interpola linealmente.
- 3) Los valores mínimo o máximo de la muestra se asignan a los cuantiles para probabilidades fuera de ese rango.

Los cuantiles más usados: **Mediana** ($q_{0.5}$), **Terciles** ($q_{0.33}$ y $q_{0.66}$),
Cuartiles ($q_{0.25}$ y $q_{0.75}$), **Quintiles**, **Deciles**, ...

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

- 1) **Localización:** Valor de “tendencia central” del conjunto
- 2) **Dispersión:** Alrededor del valor central
- 3) **Simetría:** ¿Cómo están distribuidos los datos respecto del valor central?
- 4) ...



CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

❖ Localización

En general los datos tienden a agruparse en torno a un **valor central**, el cual puede ser tomado como un **valor representativo de la muestra**.

Existen distintas formas de estimar la tendencia central de la muestra, pero los tres **estadísticos (*)** más utilizados son la moda, la mediana y la media.

(*) Un estadístico es cualquier cantidad (o función) calculada a partir de la muestra.

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

TABLA 11.3.1
Parámetros de población y estadísticas de muestra

Parámetro de la población	Estadística de la muestra
1. <i>Punto medio</i>	
Media aritmética	
$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) dx$	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Mediana	
x tal que $F(x) = 0.5$	Valor de la información en el 50o. percentil
2. <i>Variabilidad</i>	
Varianza	
$\sigma^2 = E[(x - \mu)^2]$	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Desviación estándar	
$\sigma = \{E[(x - \mu)^2]\}^{1/2}$	$s = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}$
Coefficiente de variación	
$CV = \frac{\sigma}{\mu}$	$CV = \frac{s}{\bar{x}}$

Estadístico
de la muestra

Parámetro
de la
Distribución

Es un estimador
del parámetro de
la distribución

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

TABLA 11.3.1
Parámetros de población y estadísticas de muestra

Parámetro de la población	Estadística de la muestra
1. <i>Punto medio</i>	
Media aritmética	
$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) dx$	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Mediana	
x tal que $F(x) = 0.5$	Valor de la información en el 50o. percentil
2. <i>Variabilidad</i>	
Varianza	
$\sigma^2 = E[(x - \mu)^2]$	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Desviación estándar	
$\sigma = \{E[(x - \mu)^2]\}^{1/2}$	$s = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}$

Estadístico
de la muestra

Parámetro
de la
Distribución

Es un estimador
del parámetro de
la distribución

Cuando los datos son muchos, si el ajuste de la distribución es bueno, se tiene una buena estimación de los parámetros.

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

❖ Localización

Moda: Valor más frecuente (el que aparece un mayor número de veces en la muestra).

Para variables discretas el cálculo es inmediato.

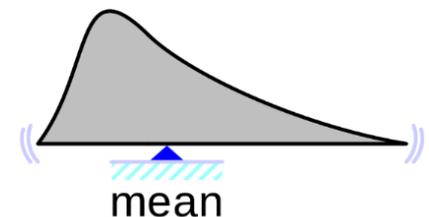
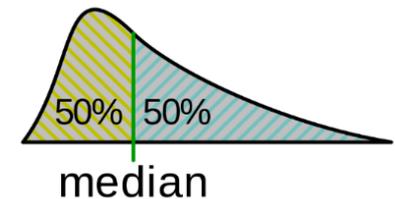
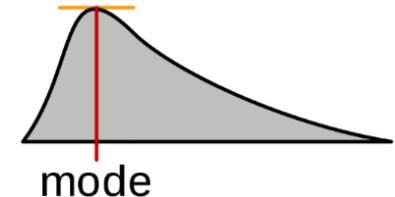
Para variables continuas puede haber más de una forma de cálculo (típicamente a partir de un histograma).

Mediana: Valor que es superado por el 50% de los datos y que supera al otro 50%.

Se calcula a partir de la muestra ordenada.

Si n es impar, la mediana es el valor del dato que queda en el lugar $(n-1)/2+1$; si n es par, la mediana será el promedio de los datos en posición $n/2$ y $n/2+1$.

Media muestral: Media aritmética de los datos.



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

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

❖ Localización

Media muestral: Media aritmética de los datos.

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

- ✓ Es el principal estadístico para caracterizar la tendencia central.
- ✓ Es un estimador de la media de la población μ .

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

❖ Localización

Media muestral: Media aritmética de los datos. $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

- ✓ Es el principal estadístico para caracterizar la tendencia central.
- ✓ Es un estimador de la media de la población μ .

Ejemplo (¡con muy pocos datos!)

2	4	<u>9</u>	11	14	$\bar{x} = 8$
2	4	<u>9</u>	11	7004	$\bar{x} = 1406$

¿Valor fuera de rango?

La media no es robusta ni resistente

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

❖ **Dispersión**

Una vez definida la tendencia central de los datos, interesa saber:

- ✓ ¿Qué tan dispersos están los datos respecto a ésta?
- ✓ ¿Cuál es la variabilidad de los datos respecto a la tendencia central?

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

❖ **Dispersión**

Existen varias medidas de la dispersión de los datos:

- ✓ **Rango**: Diferencia entre el valor máximo y el mínimo de la muestra.
- ✓ **Intervalo intercuartil** (Iqr): Diferencia entre el primer y el tercer cuartil.
- ✓ **MAD**: Desviación absoluta de la mediana, **MAD = median $\{|x_i - q_{0.5}|\}$**
- ✓ **Desviación estándar muestral**: $\hat{s} = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$
- ✓ **Coefficiente de variación muestral**: $\nu = \frac{\hat{s}}{\bar{x}}$

La más utilizada es la desviación estándar muestral.

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

❖ **Dispersión**

Desviación estándar muestral

$$\hat{s} = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

\hat{s} es un estimador de la desviación estándar de la población σ

Ni robusto ni resistente

$$\hat{s}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

\bar{s} al cuadrado es un estimador de la **varianza** de la población σ^2

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos

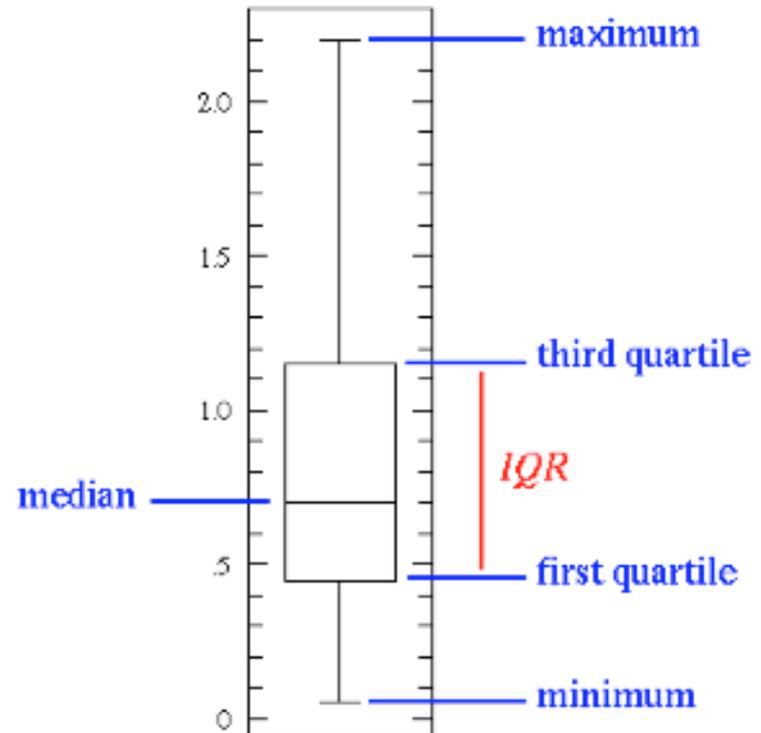
❖ **Dispersión**

Intervalo intercuartil

$$IQR = q_{0.75} - q_{0.25}$$

“No usa” el 25% superior e inferior de los datos

Robusta y resistente, pero presta poca atención a los datos extremos

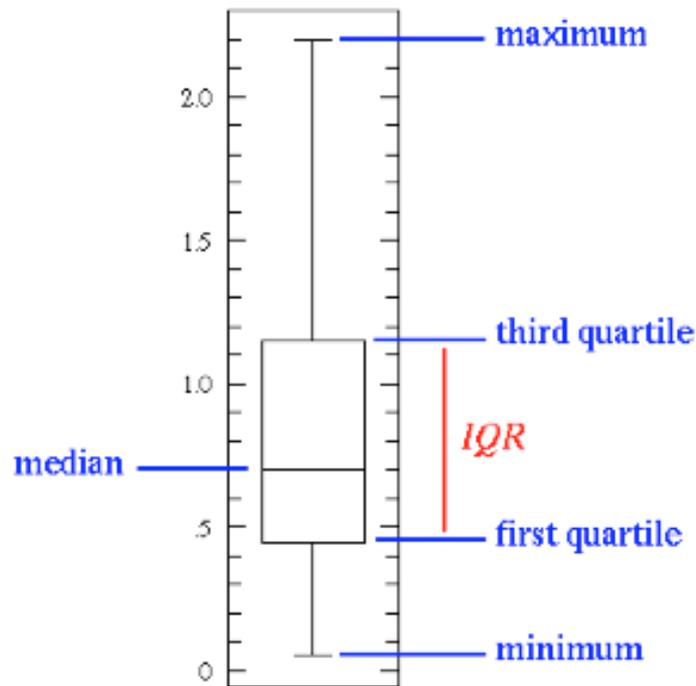


CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Principales medidas numéricas de resumen de un conjunto de datos



❖ Diagramas de cajas / Boxplots



Usa unos pocos valores para describir un conjunto de datos: la mediana, el primer y tercer cuartiles, el máximo y el mínimo y/o algún otro percentil.

CUIDADO:

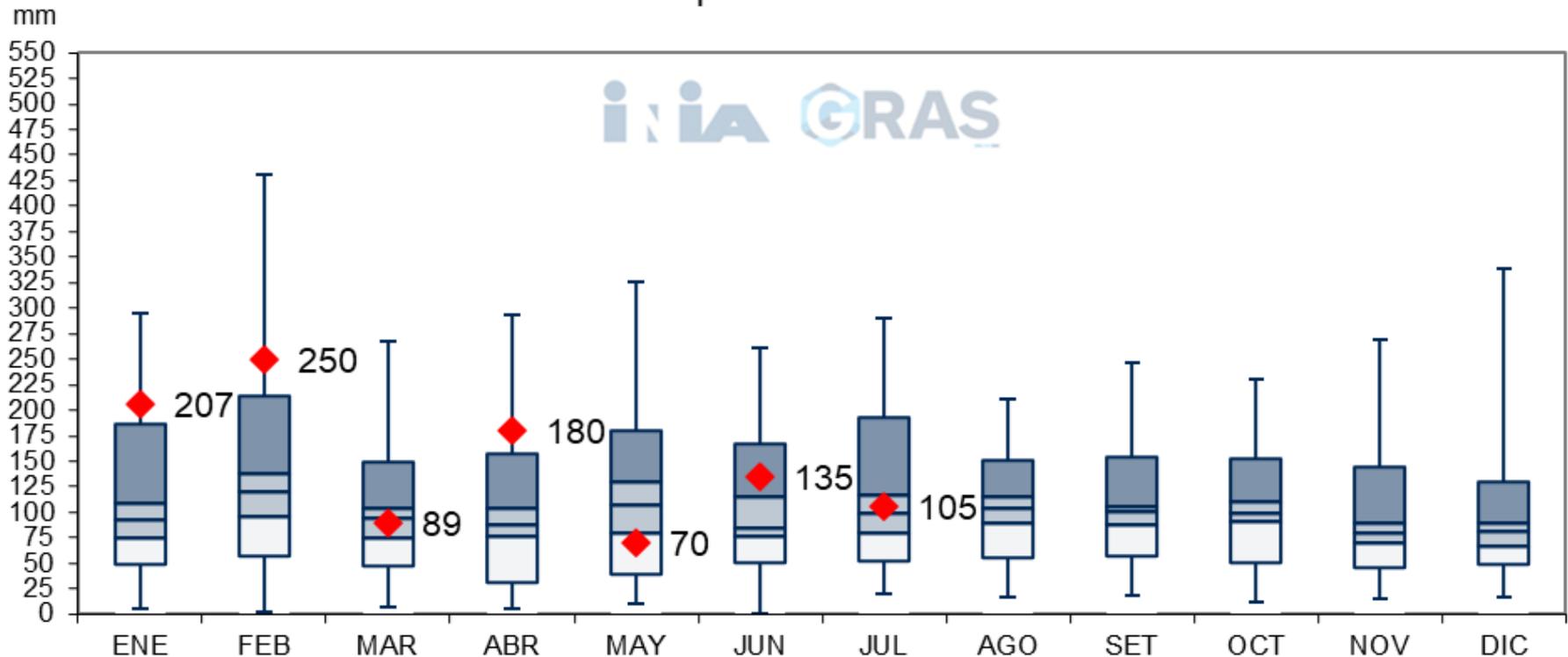
No existe una única forma de presentar los boxplots
(Ver “instrucciones” en cada caso)

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Diagramas de cajas (Boxplots)

Ejemplo: Treinta y Tres

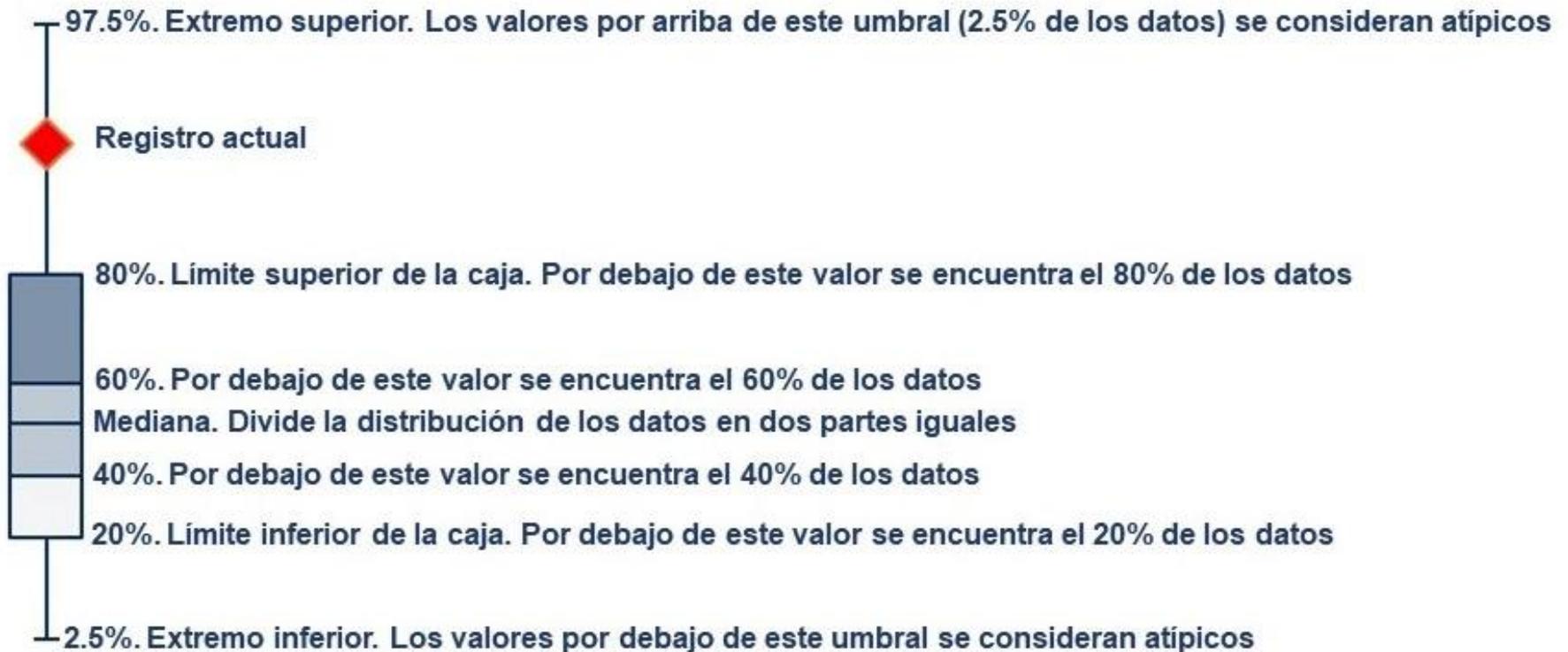
Precipitación mensual



CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Diagramas de cajas (Boxplots)

Instrucciones



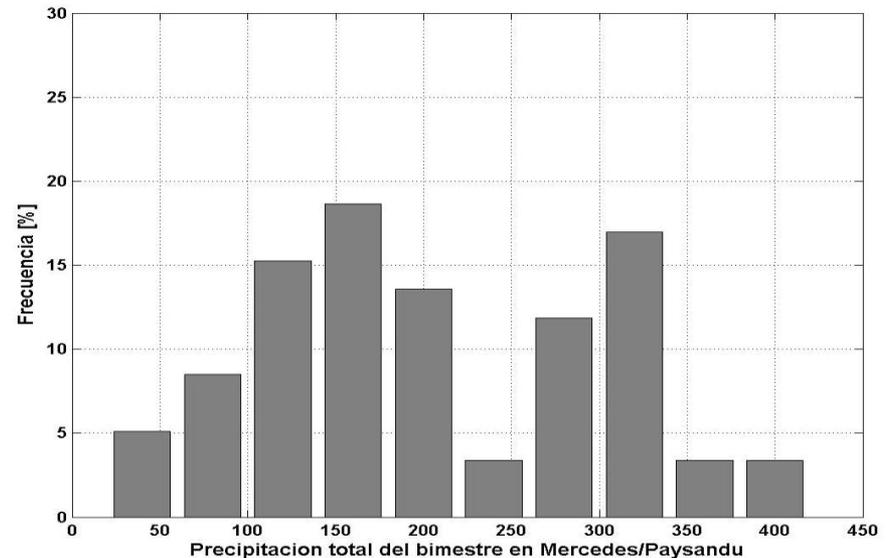
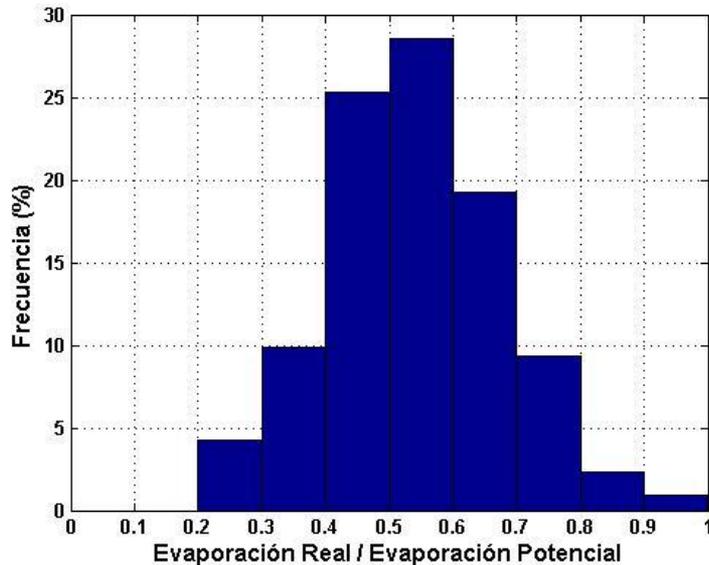
“Dato con contexto (climatología)”

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Representaciones gráficas

❖ Histograma de frecuencias

No importa el ordenamiento temporal (1 point statistics)



- ✓ ¿Cuán “poblado” queda el histograma?
- ✓ ¿Datos multimodales? (Uni-modalidad, bi-modalidad)
- ✓ ¿Huecos? ...

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Representaciones gráficas

- ❖ **Distribución empírica de frecuencia acumulada**
No importa el ordenamiento temporal (1 point statistics)

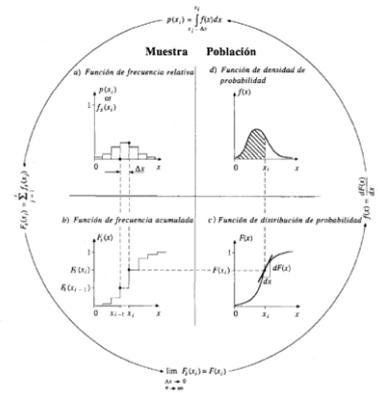
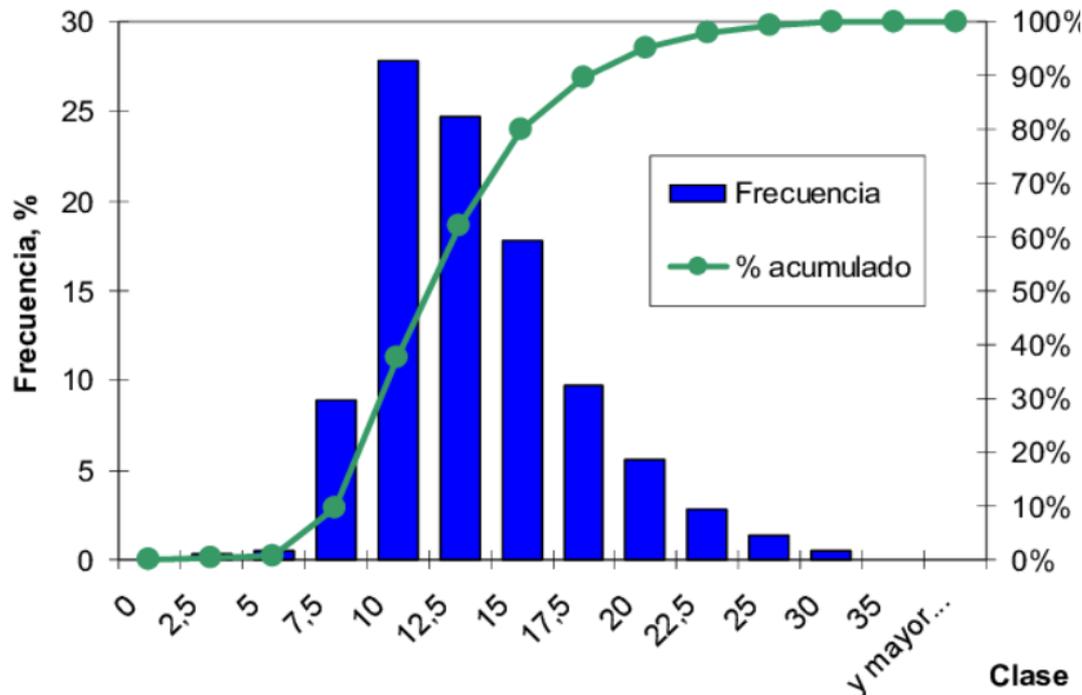


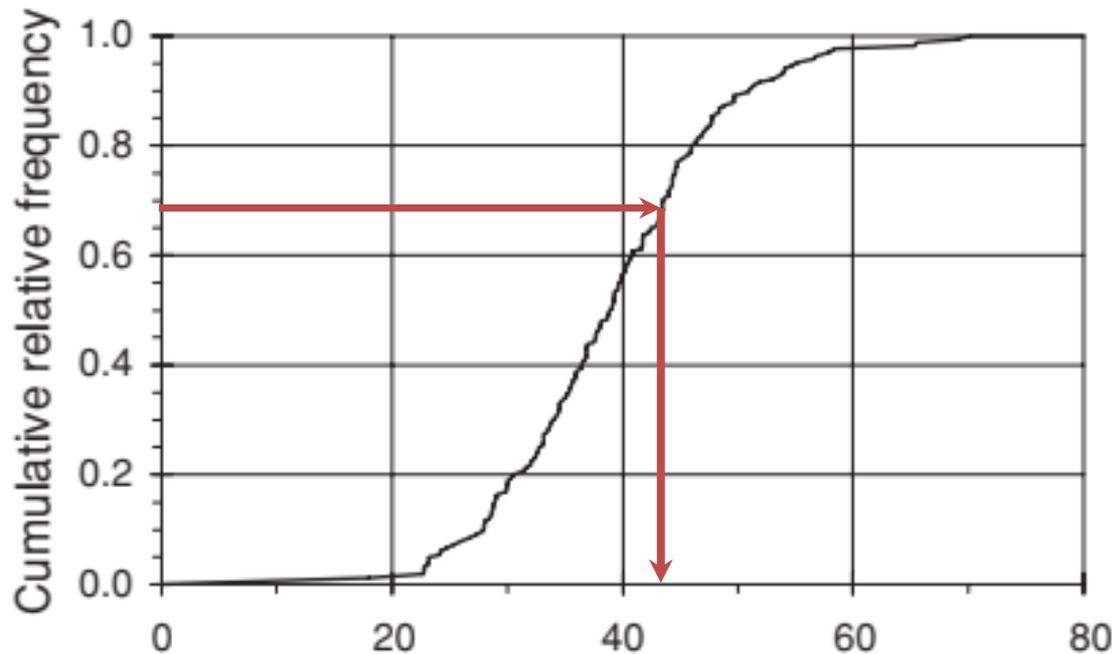
Figura 11.2.1
Ven Te Chow (1994)

Distribución empírica de frecuencia **acumulada** → Función de distribución **acumulada** (CDF)
(cuando el tamaño de la muestra tiende a infinito y el intervalo a cero)

CARACTERIZACIÓN DE LA INFORMACIÓN HIDROLÓGICA

Representaciones gráficas

- ❖ **Distribución empírica de frecuencia acumulada**
No importa el ordenamiento temporal (1 point statistics)

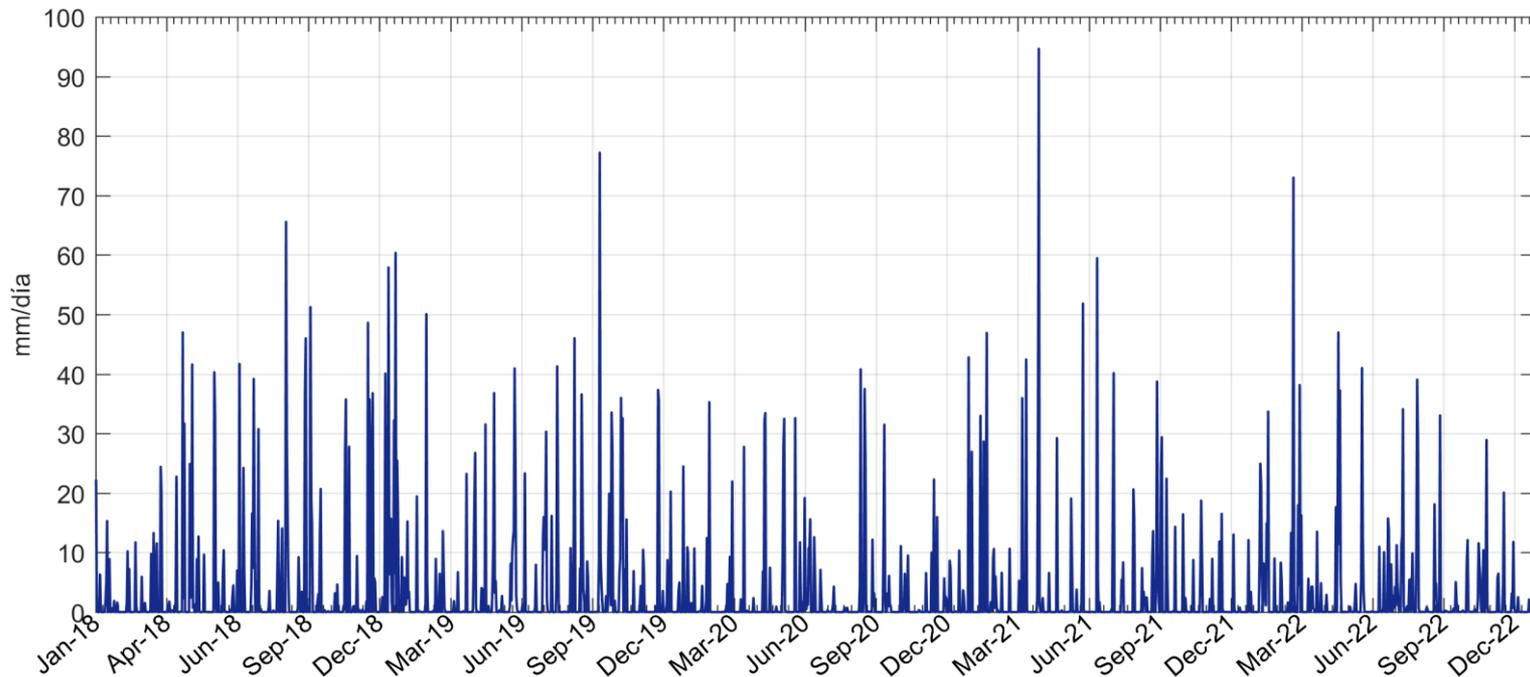


Percentiles: pth percentil es el valor que es mayor a p% de los datos

RELACIÓN ENTRE DOS VARIABLES

Una variable hidrológica y el tiempo

Gráfico: Serie temporal
(Sí importa el ordenamiento)



Precipitación diaria media areal en la cuenca del río Negro

RELACIÓN ENTRE DOS VARIABLES “APAREADAS”

Dos variables hidro-climáticas

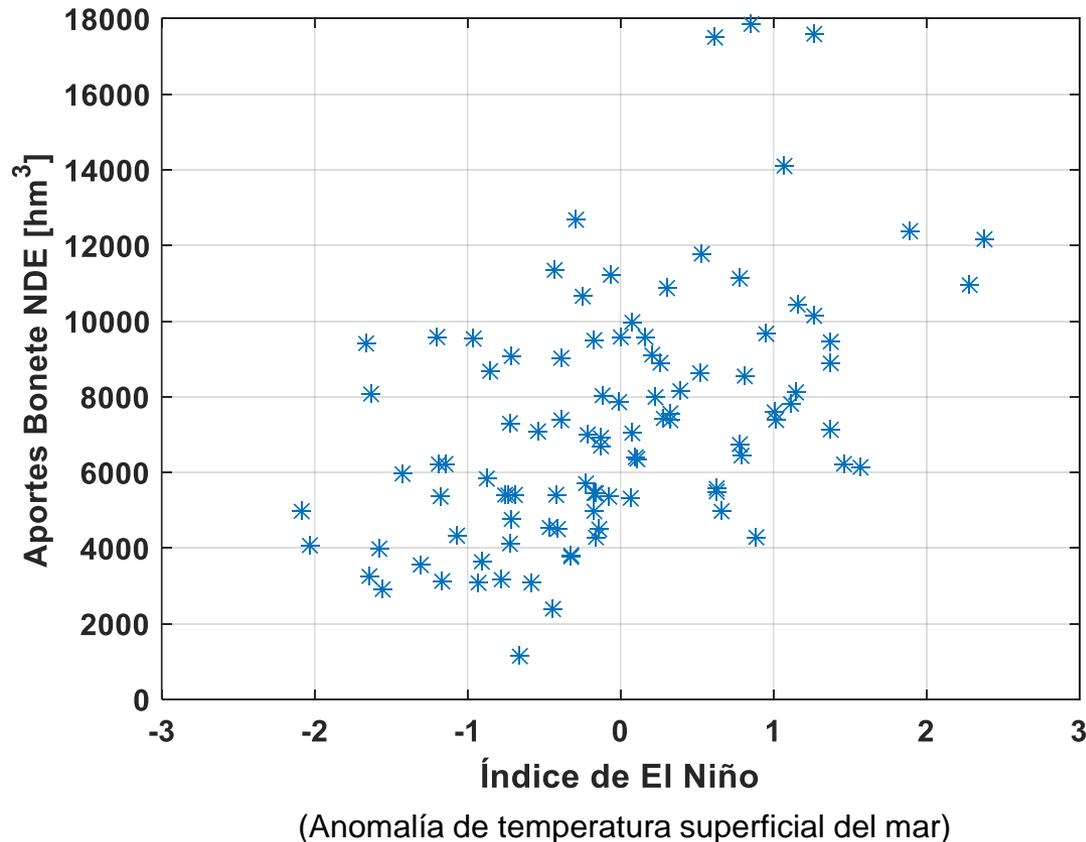
Analizar las relaciones entre dos (o más) variables hidrológicas (precipitación, temperatura, ETP, escurrimiento, etc.), en las que uno **presupone** que existe **cierta relación** y lo que buscamos es **cuantificarla**.

RELACIÓN ENTRE DOS VARIABLES “APAREADAS”

Dos variables hidro-climáticas

Gráfico de Dispersión / Scatter Plot

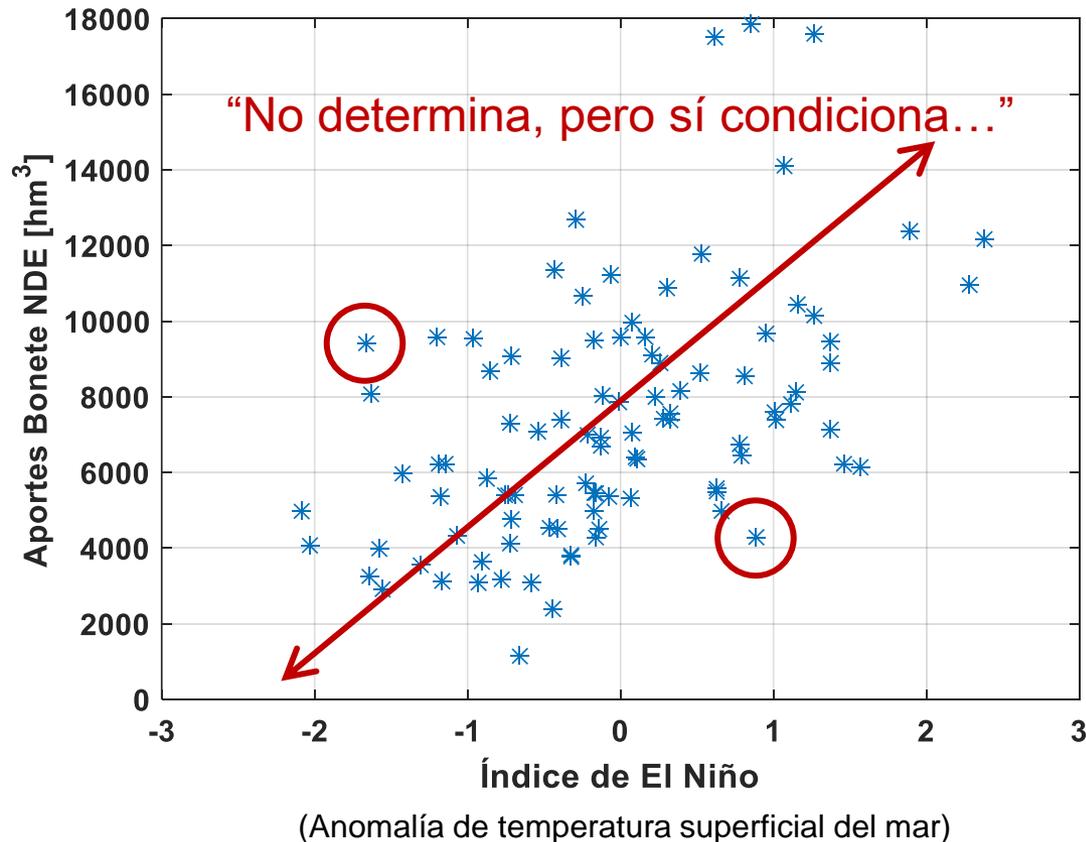
El Niño – Oscilación Sur (ENOS)
Interacción cuasi-periódica entre la atmósfera y el Océano Pacífico ecuatorial, que introduce sesgos en la distribución esperada del clima en nuestra región.



RELACIÓN ENTRE DOS VARIABLES “APAREADAS”

Dos variables hidro-climáticas

Gráfico de Dispersión / Scatter Plot



REGRESIÓN LINEAL SIMPLE

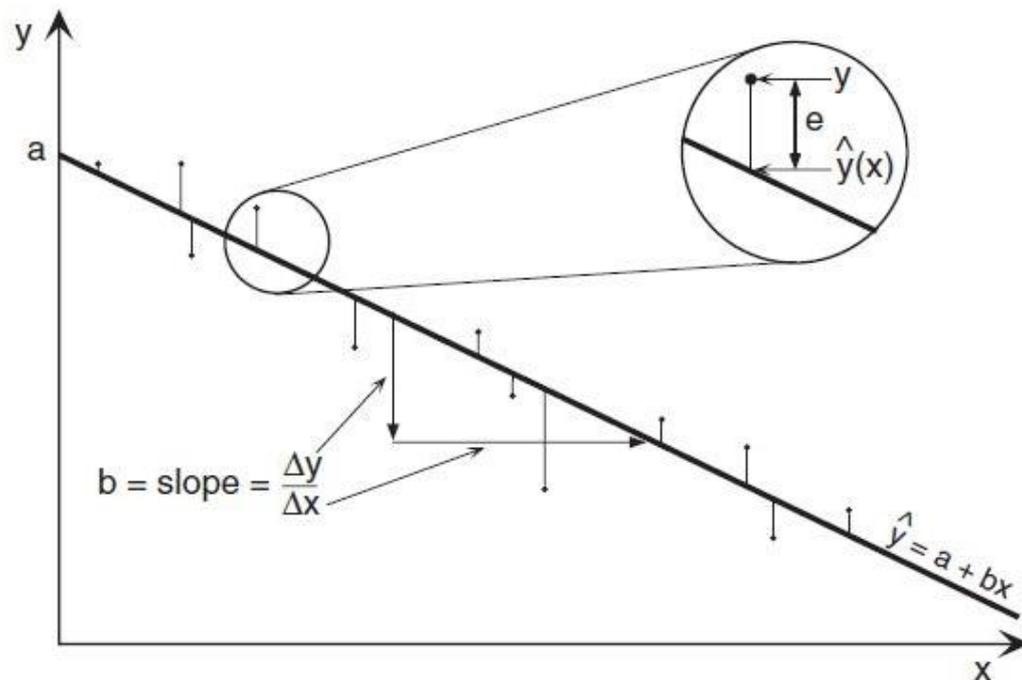
Estimación por Mínimos Cuadrados

❖ Mínimo error cuadrático: $\sum e_i^2$

$$\hat{y} = a + b \cdot x$$

$$y_i = a + b \cdot x_i + e_i$$

$$e_i = y_i - \hat{y}(x_i)$$



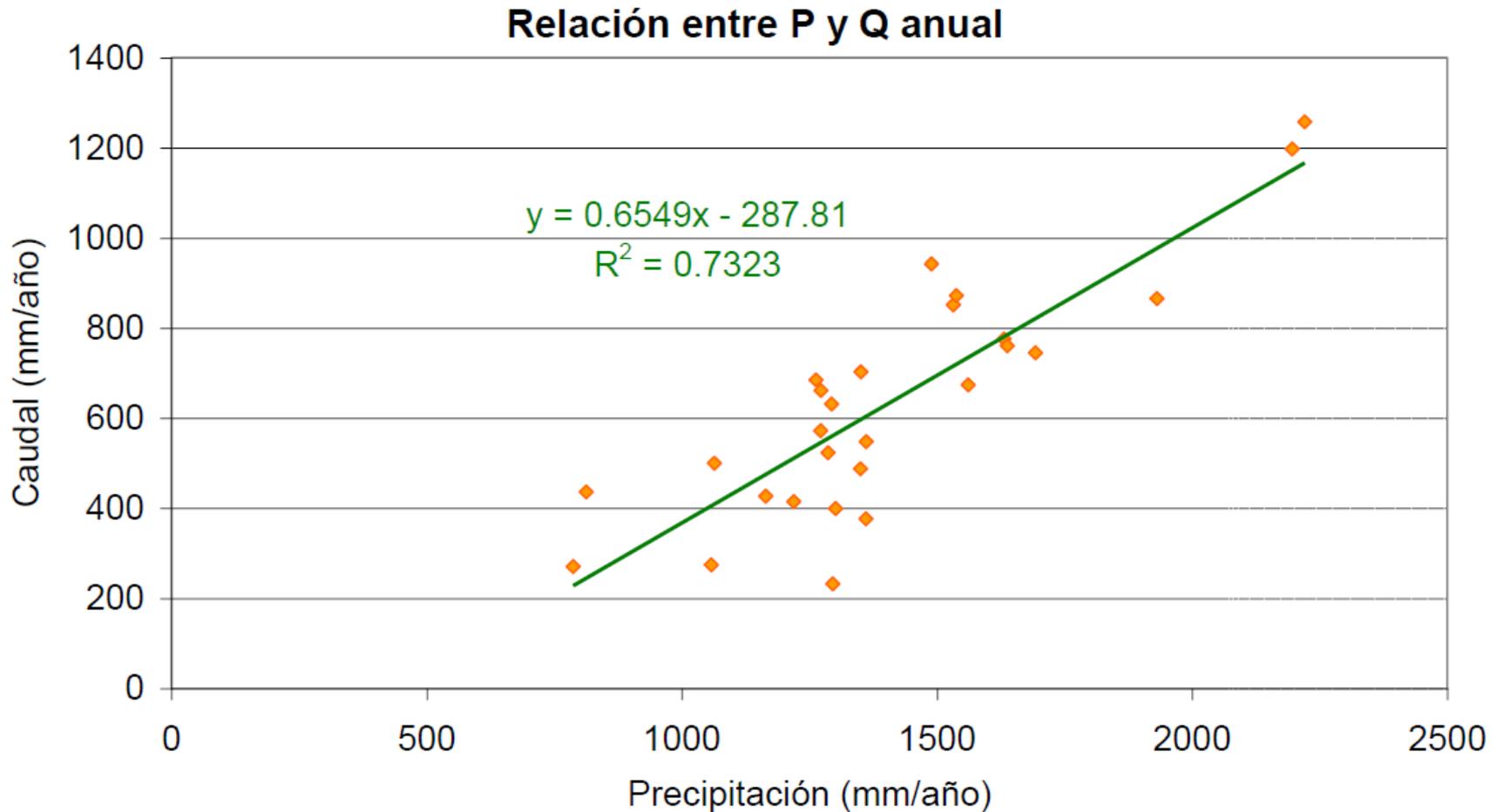
Solución:

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

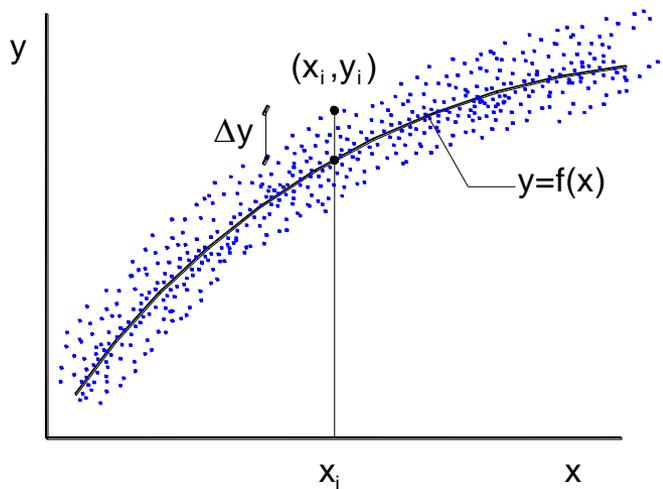
REGRESIÓN LINEAL SIMPLE

Ejemplo



COEFICIENTE DE CORRELACIÓN DE PEARSON

Coeficiente adimensional que proporciona una medida de cuán fuerte es la correlación **lineal** entre dos variables.



$$r = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$$

$$Cov(x, y) = \frac{1}{n - 1} \sum (x - \bar{x}) \cdot (y - \bar{y})$$

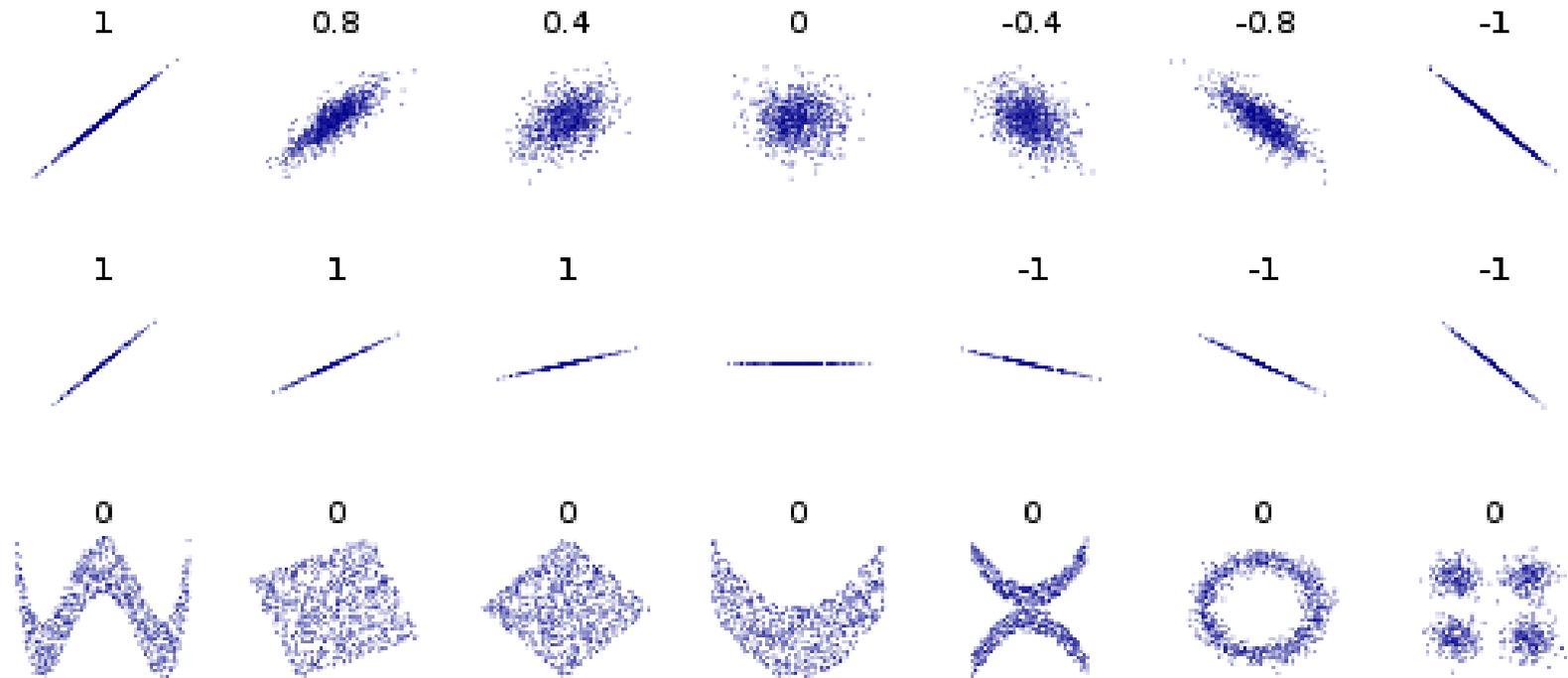
$$-1 < r < 1$$

- ❖ Cuando $r \rightarrow 0$: las variables son no correlacionadas
- ❖ Cuando $|r| \rightarrow 1$: las variables están fuertemente correlacionadas
- ❖ La correlación no evidencia una relación causal

Coeficiente de determinación: $R^2 = r^2$

COEFICIENTE DE CORRELACIÓN DE PEARSON

$$-1 \leq r_{xy} \leq 1$$



MODELOS DE REGRESIÓN LINEAL

Recomendación

La facilidad para estimar los parámetros conduce a la tentación de aplicar el procedimiento irreflexivamente...

- ❖ ¿Existe un argumento físico para seleccionar la variable explicativa x para estimar la variable dependiente y ?
- ❖ ¿La relación parece ser lineal o curva? ¿Cambia la variabilidad de y para diferentes rangos de valor de x ?
- ❖ ¿El coeficiente de correlación es “significativamente distinto” de cero? (Test de significancia estadística (*))
- ❖ Análisis de los residuos

(*) Verificación de que un enunciado no se pueda explicar simplemente por el azar.
Es crítico cuando las muestras son pequeñas.

CONTROL DE CALIDAD DE DATOS



En ausencia de **contexto** alguno es muy difícil detectar errores puntuales o sistemáticos que no sean elementales



Sin **contexto**, solo se puede dar alertas de datos sospechosos que se deberán verificar.



El **control de calidad de datos** es una disciplina en sí misma, daremos sólo algunos elementos de guía.

CONTROL DE CALIDAD DE DATOS

Contextos que posibilitan la detección de errores

❖ Errores puntuales en datos aislados

- ✓ Estaciones de medición “suficientemente” cercanas
- ✓ Otra información relevante (lluvia - escurrimiento)

❖ Errores sistemáticos en períodos de datos

- ✓ Conocimiento de estadísticos de la distribución de la variable (número de días de lluvia, precipitación media, ciclo anual, etc.)
- ✓ Estaciones de medición “suficientemente” cercanas
- ✓ En ausencia de conocimiento de los estadísticos o de acceso a datos cercanos: hipótesis de estacionariedad de la serie puede detectar errores en base a tendencias de estadísticos (p.e. media)

CONTROL DE CALIDAD DE DATOS

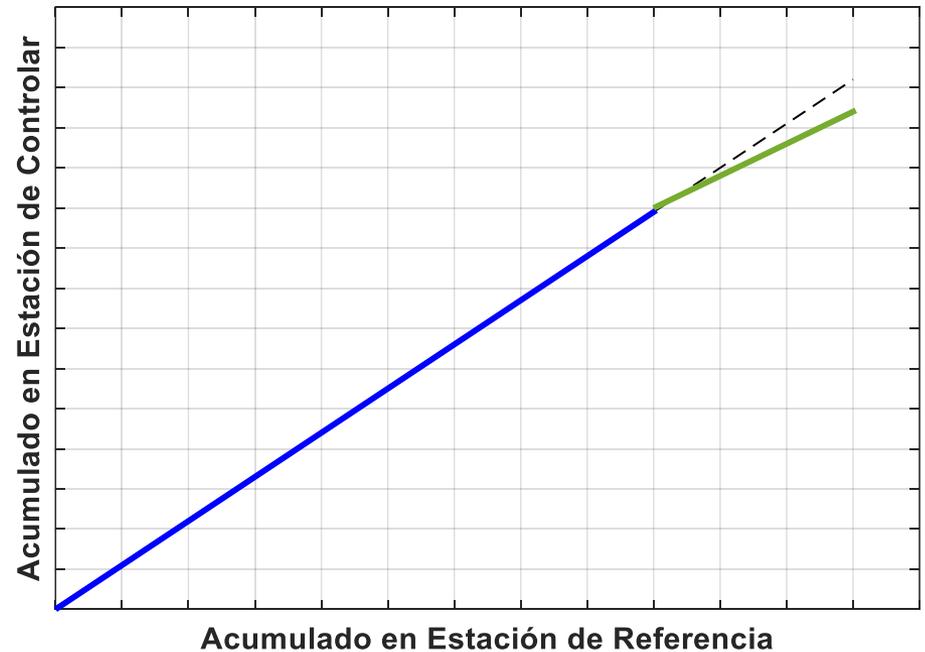
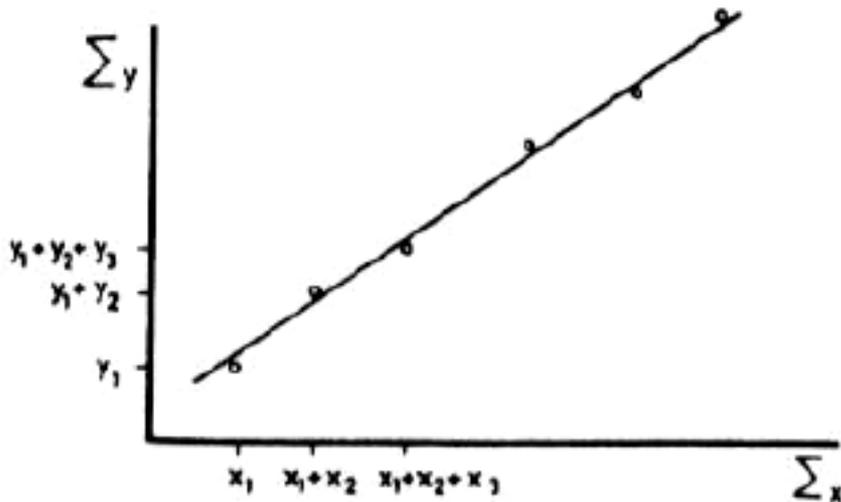
En contexto del país y estaciones vecinas

- ❖ Verificar **estadísticos robustos**:
 - ✓ Frecuencia media de lluvia (¿menores a 20%?)
 - ✓ Media anual (¿muy baja/alta?)
- ❖ Verificar **lluvias excepcionales** (¿ $P > 100$ mm/día?)
- ❖ Acumulado de un año demasiado lejano a **estación contigua**, ¿cuál es más confiable?
- ❖ Métodos más objetivos para errores sistemáticos:
 - ✓ Doble masa (Gilman, 1964)
 - ✓ Análisis de Componentes Principales

CONTROL DE CALIDAD DE DATOS

En contexto del país y estaciones vecinas

- ❖ **Método de Doble Masa**
(Consistencia de datos)

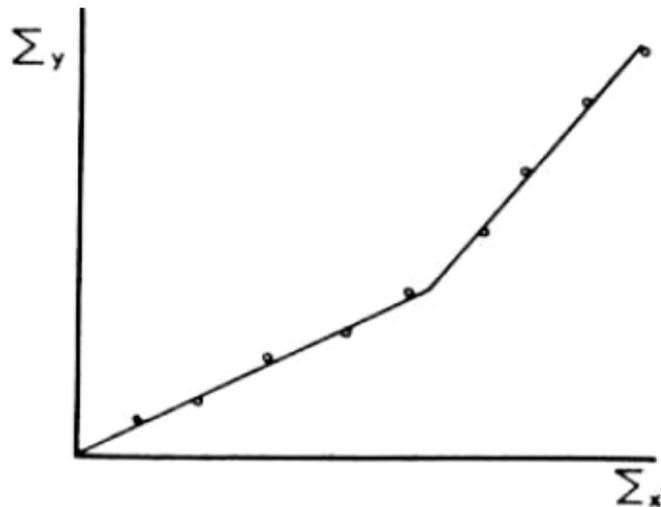


CONTROL DE CALIDAD DE DATOS

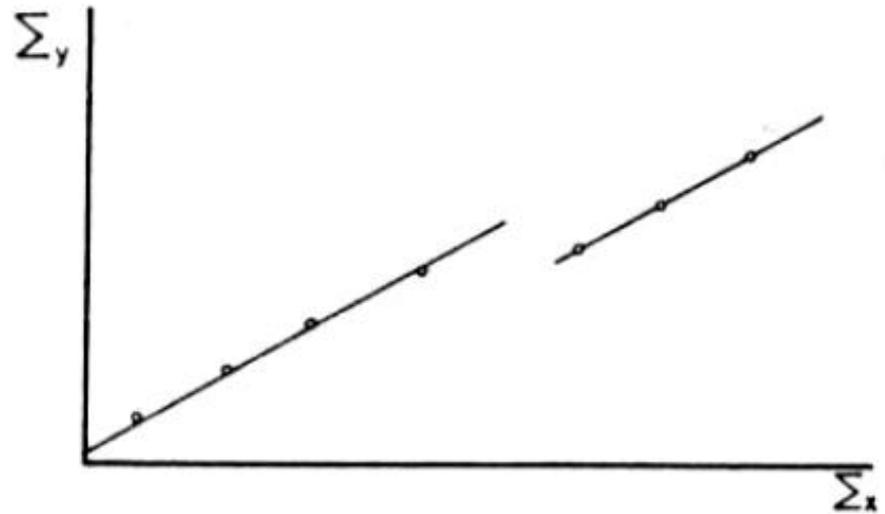
En contexto del país y estaciones vecinas

❖ Método de Doble Masa

Quiebres, saltos, ...



¿Cambio de sitio?



¿Mal funcionamiento en un año?

CONTROL DE CALIDAD DE DATOS

Valores fuera de rango (Outliers)

Se denomina **outlier** a datos con **valores inesperadamente altos o bajos**, que no parecieran seguir la misma distribución que el resto de los datos.

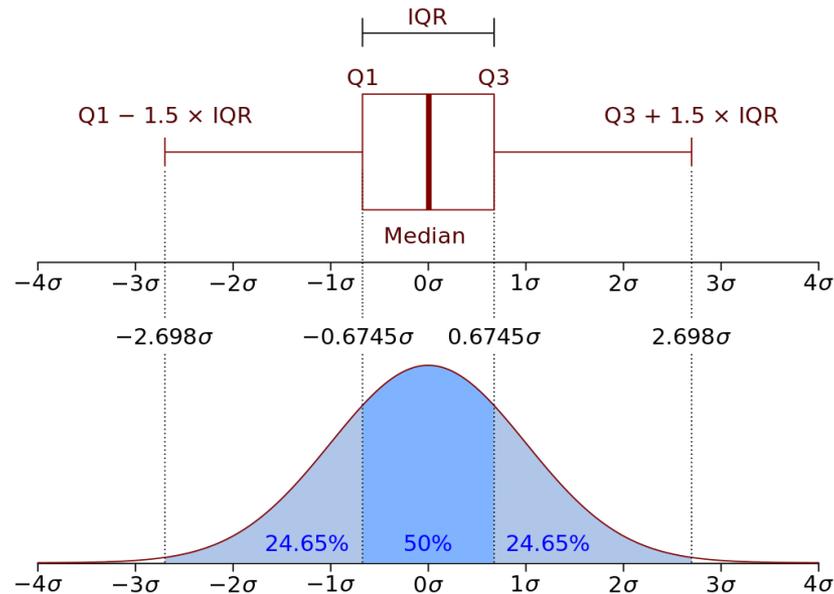
- ❖ Para pasar de outlier a “candidato al descarte o corrección” se requiere **contexto** (climatología, estaciones cercanas, ...).
- ❖ Estos valores pueden estar asociados a **errores de medida o de lectura** y, en estos casos, deberán ser descartados del análisis.
- ❖ Pero también pueden estar justificados o tener una **explicación física**. En este caso habrá que decidir cómo se utiliza la información proporcionada por estos datos según el tipo de análisis estadístico que se vaya a hacer.

CONTROL DE CALIDAD DE DATOS

Valores fuera de rango (Outliers)

En algunos casos se incluyen **reglas de identificación de outliers**:

- ❖ Por ejemplo: todo dato menor a $Q1 - 1.5 \text{ iqr}$ o mayor a $Q3 + 1.5 \text{ iqr}$



CAUDALES



En general surge de la **observación de niveles** en secciones aforadas.



Los aforos tienen **rangos de validez**.



Medida puntual que **integra procesos hidrológicos** en la cuenca de aporte.



Para analizar la relación P-Q también se incorpora la incertidumbre de la estimación del campo de precipitación.

CAUDALES

Curva de permanencia

Representa la frecuencia con que ocurren valores iguales o superiores a los valores de una serie temporal.

- ❖ Es muy utilizada para evaluar el potencial de una sección fluvial (¿cuánto caudal circula con X% certeza?).
- ❖ Determinar la garantía de contar con caudales iguales o mayores a la demanda que se desea abastecer durante un (X%) del tiempo.

CURVA DE PERMANENCIA

Construcción

A partir de la serie de datos de un paso de tiempo dado:

- a) Ordenar la variable temporal en **orden decreciente**, atribuyendo orden 1 al mayor valor y orden n al menor valor, en una muestra de tamaño n .
- b) Calcular la **frecuencia** con que cada valor ordenado es excedido o igualado (permanencia), como $100 \cdot (m/n)$, siendo m el orden y n el tamaño de la muestra.
- c) Graficar la serie ordenada con la escala de permanencia representada en el eje de las abscisas.

Es esencialmente una función de frecuencia acumulada,
contiene exactamente la misma información.

CURVA DE PERMANENCIA

Construcción

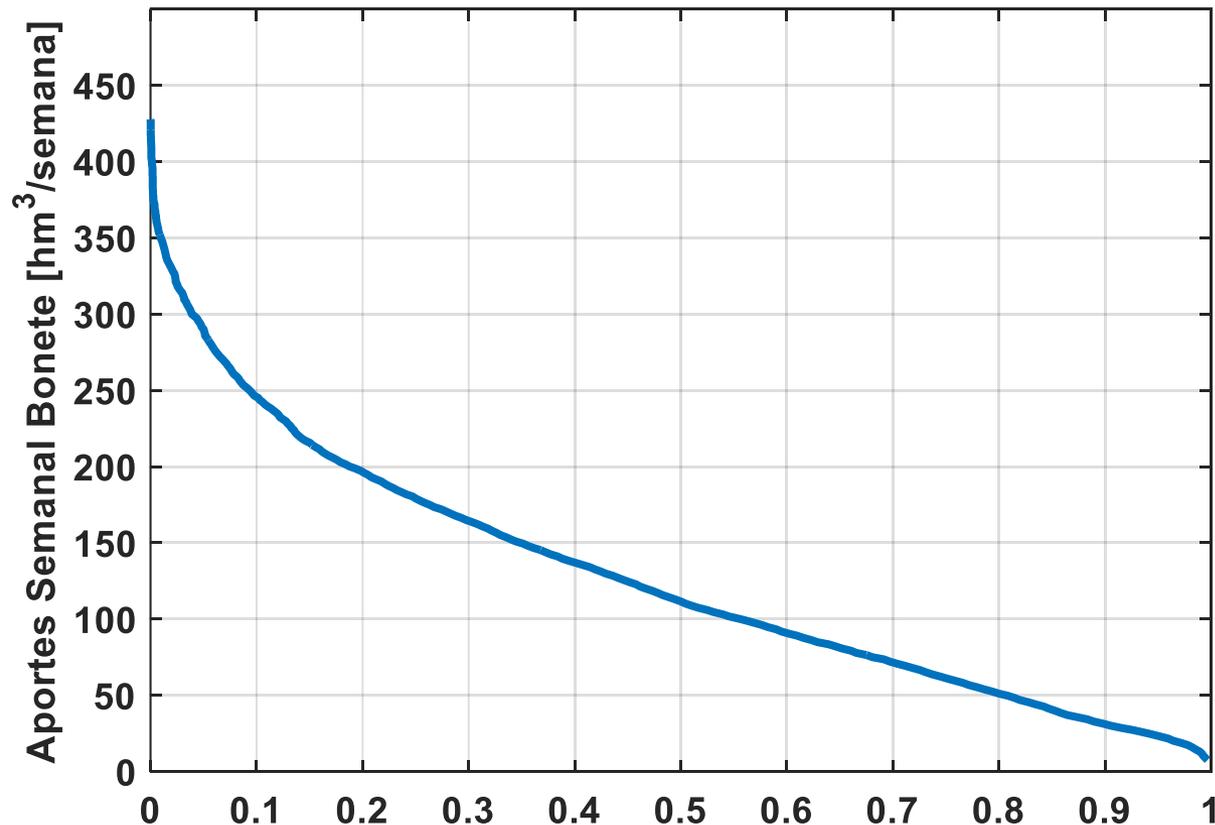
100%
42 datos

Orden	Permanencia (%)	Q (m3/s)
1	2,38	17,622
2	4,76	15,711
3	7,14	14,491
4	9,52	13,471
5	11,90	13,251
6	14,29	13,047
7	16,67	12,945
8	19,05	12,719
9	21,43	12,496
10	23,81	12,226
11	26,19	11,993
12	28,57	11,932
13	30,95	11,697
14	33,33	11,376
15	35,71	11,328
16	38,10	11,307
17	40,48	10,661
18	42,86	10,470
19	45,24	10,231
20	47,62	9,841
21	50,00	9,281

Orden	Permanencia (%)	Q (m3/s)
22	52,38	8,647
23	54,76	8,389
24	57,14	8,336
25	59,52	8,315
26	61,90	7,529
27	64,29	7,410
28	66,67	7,102
29	69,05	6,625
30	71,43	6,513
31	73,81	6,439
32	76,19	5,950
33	78,57	5,805
34	80,95	5,598
35	83,33	5,553
36	85,71	4,889
37	88,10	4,866
38	90,48	4,527
39	92,86	4,186
40	95,24	3,842
41	97,62	3,037
42	100,00	2,912

CURVA DE PERMANENCIA

Curva de permanencia de caudales SEMANALES en Bonete



¡ 101 años de datos !

Ojo con representatividad del período de muestreo sobre todo para los extremos

HIDROLOGÍA ESTADÍSTICA



Edición 2025

Alejandra De Vera (en base a notas de Rafael Terra)

Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA)
Facultad de Ingeniería, Universidad de la República, Uruguay

adevera@fing.edu.uy