ESTADISTICA DE EXTREMOS

Profesor Luis F. Carvajal

Universidad Nacional de Colombia

Facultad de Minas

Escuela de Geociencias y Medio Ambiente

Datos

1. Según el propósito de diseño extraer del registro el tipo de datos indicado (máximos instantáneos, máximos diarios, etc.)
2. Hay dos formas de extraer la muestra:

a. Tomar el valor máximo de cada año (máximos anuales). Intervalo de recurrencia: número promedio de años entre la ocurrencia de un evento de cierta magnitud.
b. Tomar los n valores máximos entre los n años de registro. Los eventos deben ser independientes (excedencias anuales). El intervalo de recurrencia no tiene implicación.

Datos

Cuando usar una muestra de extremos anuales o la de excedencias anuales.

a. Cuando valores máximos seguidos se deben considerar en el análisis .(Excedencias anuales)
b. Análisis con intervalos de recurrencia bajos (Excedencias anuales)
c. Si el análisis es de eventos extremos (máximos anuales).

Homogeneidad de los datos

Se debe verificar sin la muestra seleccionada proviene de la misma distribución.

a. Establecer la hipótesis nula (Ho) de que no hay diferencia entre los dos grupos de datos.
b. Seleccionar un test estadístico.
c. Seleccionar un nivel de significancia.
d. Asumir que la distribución de muestreo del test cumple Ho.
e. Teniendo en cuenta b, c y d definir la región de rechazo.
f. Calcular el valor del estadístico. Si esta en la región de rechazo, rechazar Ho.

Test

Hay dos tipos de test, paramétricos y no paramétricos, los paramétricos deben cumplir con las siguientes consideraciones.

a. Las observaciones deben ser independientes.
b. Las observaciones deben provenir de una distribución normal.
c. Las poblaciones deben tener la misma varianza.

Prueba paramétrica: T.

Pruebas no paramétricas: Mann-Whitney, Kruskal-Wallis, test de Kolmogorov-Smirnov.

Pruebas de independencia

1. Pruebas no paramétricas: punto de cambio, corridas y de Spearman.
2. Pruebas basadas en el correlograma: prueba de Anderson.

- Prueba de punto de cambio:

$$U = \left| \frac{M - \frac{2}{3}(N-2)}{\sqrt{\frac{16N-29}{90}}} \right|$$

$U \le Z_{1-\alpha/2}$ $\quad$ Z distribución normal

M: número de picos y valles.

$$Pico \quad X_{t-1} < X > X_{t+1}$$

$$Valle \quad X_{t-1} > X < X_{t+1}$$

- Prueba de corridas:

$$U = \left| \frac{k - \bar{k}}{S_k} \right|$$

$$U \leq Z_{1-\alpha/2}$$

K es el número de corridas. Una corrida está representada por valores consecutivos de 0s y 1s en la serie Zt definida de la siguiente forma:

$$Z_t = 1 \quad si \ X_t \geq \overline{X}$$

$$Z_t = 0 \quad si \ X_t < \overline{X}$$

$$n_1 = \sum_{t=1}^{N} Z_t$$

$$n_2 = 1 - \sum_{t=1}^{N} Z_t$$

$$n_1 + n_2 = N$$

$$\overline{K} = \frac{2 n_1 n_2}{n_1 + n_2} + 1$$

$$S_k = \left[\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} \right]^{1/2}$$

- Prueba del coeficiente de Spearman.

$$U = \left| \frac{R\sqrt{N-2}}{\sqrt{1-R^2}} \right|$$

$$R = 1 - \frac{6\sum_{t=1}^{N} D_t^2}{N\left(N^2-1\right)}$$

$$U < t_{1-\alpha, N-2}$$

R es el coeficiente de Spearman. t es el cuantil de la distribución t de Student. Dt la diferencia entre la posiciones de un valor Xt entre la serie original y ordenada de menor a mayor.

- Prueba de Anderson:

Se estima la banda de confianza para el correlograma de la muestra. Si el correlograma cae entre las bandas, los datos de la muestra son independientes.

Para el cálculo de lo límites se usa la siguiente expresión:

$$\frac{-1 \pm Z_{1-\alpha/2}\sqrt{N-k-1}}{N-k}$$

Z es la distribución normal estandarizada.

Introducción

1. Los eventos hidrológicos son COMPLEJOS, ERRATICOS, y de NATURALEZA ALEATORIA.
2. Por la escala espacio-temporal es difícil su modelación física.
3. La Estadística en la Hidrología es básica para el diseño en la ingeniería (definición de condiciones críticas).
4. La definición del régimen hidrológico requiere: Análisis estadístico y probabilístico basados en registros históricos.
5. La Hidrología trata con VARIABLES ALEATORIAS.

Eventos hidrológicos

Variables

VARIABLE ALEATORIA (v.a): Variable cuyo comportamiento no puede predecirse con certidumbre.

- **v.a Discreta:** Solo puede tomar valores específicos. La ley de probabilidades asocia medidas de probabilidad a cada posible ocurrencia de la v.a.
- **v.a Continua:** Si puede tomar todos los valores en un rango de ocurrencia. La ley de probabilidades asocia medidas de probabilidad a rangos de ocurrencia de la v.a.

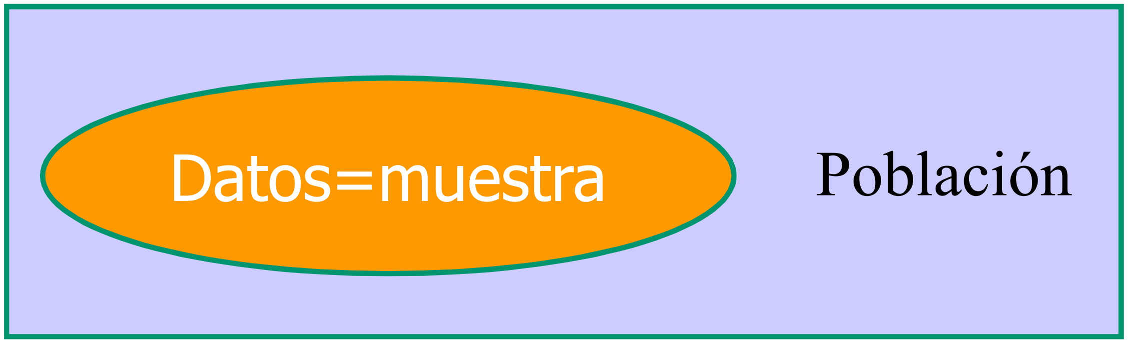
Definiciones

1. **Espacio muestral:** el conjunto de todas las posibilidades. Los espacios muestrales pueden ser discretos o continuos.

2. **Punto muestral:** cada una de las posibilidades.

3. **Evento:** subconjunto del espacio muestral.

La información histórica de una variable hidrológica representa una MUESTRA de la POBLACIÓN.



1. **ANÁLISIS PROBABILÍSTICO:** Análisis de POSIBLES Leyes de Probabilidad que describan comportamiento de la población.
2. **ANÁLISIS ESTADÍSTICO:** Se hacen inferencias sobre la variable (Población) usando la MUESTRA

Problema importante

Interpretación de registros históricos INFERIR la Ley de Probabilidades de la Variable hidrológica de interés:

ANALISIS DE FRECUENCIA

Tipos de análisis

- **DESCRIPTIVO:** Este se realiza sin ninguna referencia a su POBLACIÓN. Se calculan propiedades estadísticas (aplicación métodos estadísticos). Poco RIESGO.
- **INFERENCIA:** Se analiza la muestra para inferir las propiedades de la población, lo cual ayuda a derivar características probabilísticas. Implica RIESGO.

Estadística en Hidrología

1. Interpretación de observaciones
2. Análisis de la calidad de la información
3. Inferencia sobre el comportamiento de la variable
4. Extracción del máximo de información de los registros
5. Información en gráficas tablas, ecuaciones para toma de decisiones en el planeamiento en R.H.

Conceptos de Probabilidad

1. Métodos estadísticos: Basados en principios matemáticos que describen la variación aleatoria de un conjunto de observaciones de un proceso. Importan los datos en sí mas que el proceso físico que los produjo.

2. Una variable aleatoria X está descrita por una distribución de probabilidad. La distribución especifica la probabilidad de que una observación x de la variable, esté en un rango especificado de X.

3. Si una muestra de n observaciones tiene n_i valores en el rango del evento x, entonces la frecuencia relativa es n_i/n

Axiomas

Para un evento E en un espacio muestral S:

$$P(E) \geq 0$$

$$P(S) = 1.0$$

$$P(E_1 U E_2) = P(E_1) + P(E_2)$$

Eventos independientes: Si la probabilidad de ocurrencia de uno no se ve afectada por la ocurrencia del otro.

Eventos mutuamente excluyentes: Cuando la ocurrencia de uno imposibilita la ocurrencia del otro.

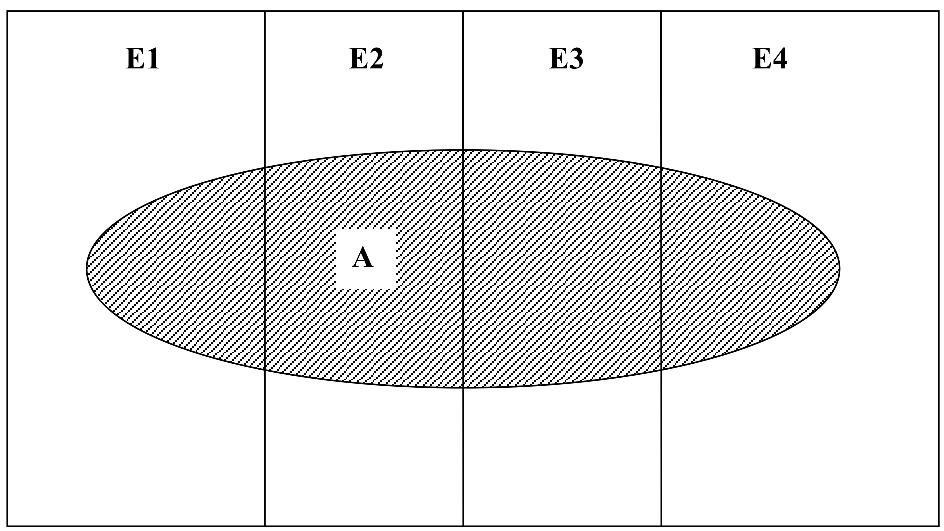
Principios de la Probabilidad

1. Probabilidad Total

Espacio muestral Ω dividido en ***m*** eventos, A_1, A_2, … , A_m, entonces:

$$P(A_1) + P(A_2) + ... + P(A_m) = 1$$

Teorema de la Probabilidad Total



P(A) = P(A/E1)+ P(A/E2)+P(A/E3)+………P(A/E_i)

2. Complementariedad

Si $\overline{A}$ es el complemento de A, entonces: $\overline{A} = \Omega - A$

$$P(\overline{A}) = 1 - P(A)$$

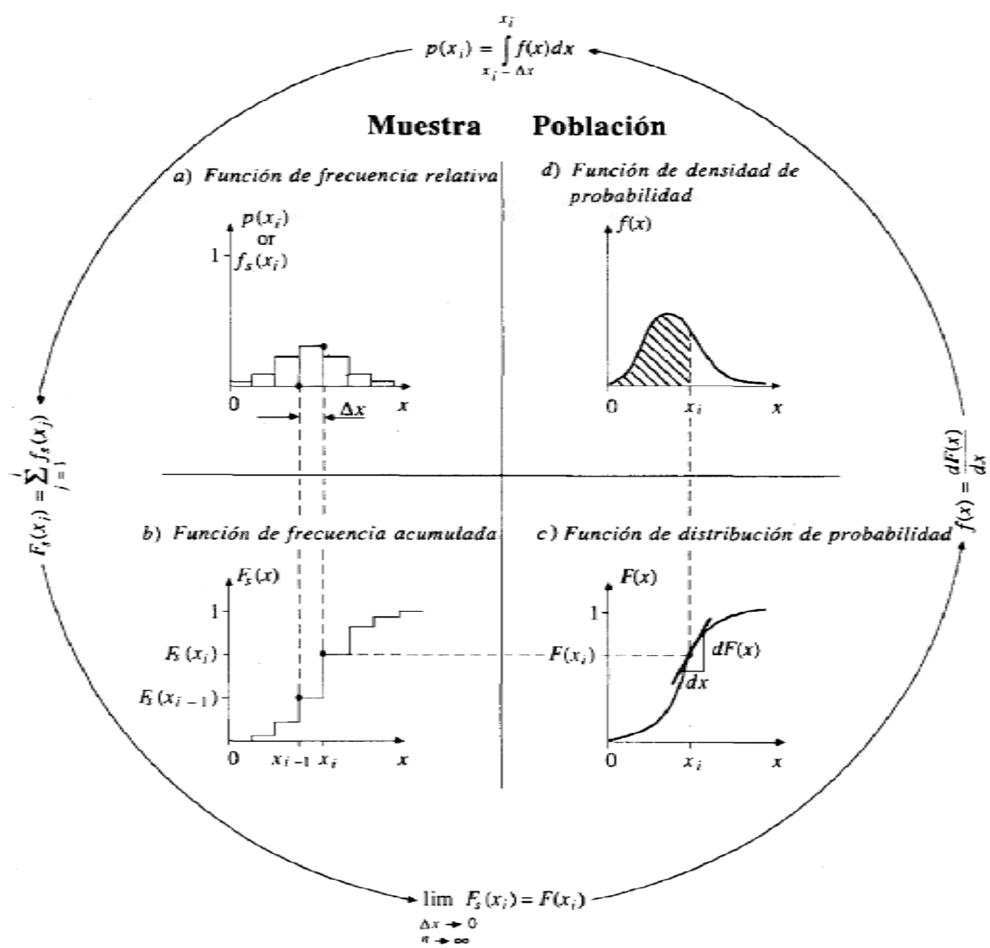
3. Probabilidad Condicional

$$P(B / A) = \frac{P(A \cap B)}{P(A)}$$

Si A y B son independientes: $P(B / A) = P(B)$

$$P(A \cap B) = P(A)P(B)$$

Funciones de probabilidad



Tomado de Hidrología Aplicada, Ven te Chow et al, pag 367.

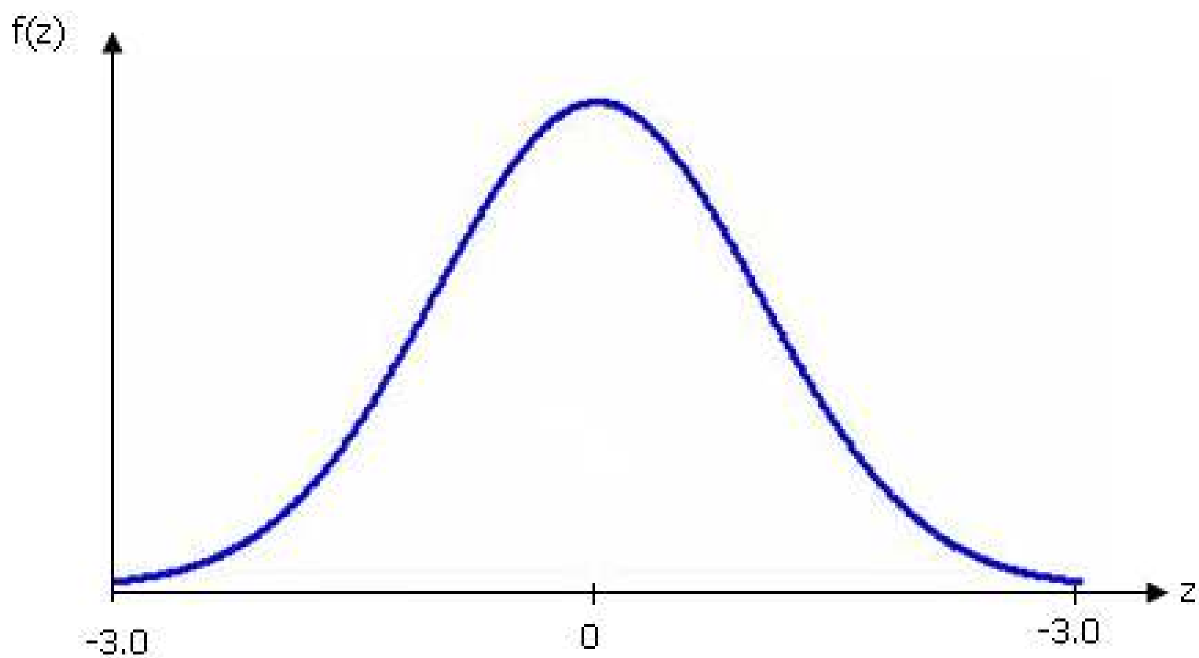
Distribución Normal

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

Variable normal estandarizada $z = \frac{(x-\mu)}{\sigma}$

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

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



Secuencia de Bernoulli

Ocurrencia de un evento o recurrencia de este en una secuencia repetida de ensayos.

Q_{max} en una serie de caudales, para el diseño.

Este tipo de problemas tienen dos resultados: ocurrencia y no ocurrencia de un evento.

La secuencia de Bernoulli:

1. Cada ensayo tiene dos resultados: falla o éxito.
2. La probabilidad de ocurrencia de un evento en cada ensayo es constante.
3. Los ensayos son estadísticamente independientes.

Distribución Binomial

Si la probabilidad de éxito u ocurrencia es p y (1 – p) es la probabilidad de falla o no ocurrencia, entonces la probabilidad de ocurrencia entre n ensayos es una secuencia de Bernoulli y está dada por:

$$P(X = x) = \binom{n}{x} P^x (1 - P)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

Donde:

$$\binom{n}{x} = \frac{n!}{[x!(n-x)!]} \longrightarrow \text{Coeficiente binomial}$$

Distribución Geométrica

En una secuencia de Bernoulli, el número de ensayos hasta que un evento específico ocurra por primera vez es modelado por la distribuciones geométrica.

Éxito $\Rightarrow$ ensayo t_{th}

Falla $\Rightarrow$ en los t – 1 ensayos

Si T $\Rightarrow$ v.a. apropiada:

$$P(T=t) = pq^{T-1} \qquad\qquad t = 1, 2, \ldots.$$

Periodo de retorno $\Rightarrow$ Tiempo de recurrencia promedio.

Periodo de retorno

- **Período de retorno:** Es el tiempo entre ocurrencias de $X \geq x_T$. El periodo de retorno (T) del evento $X \geq x_T$ es el valor esperado de T, E(T), su valor promedio medido sobre un gran número de ocurrencias.

- **Periodo de retorno:** Intervalo promedio de recurrencia entre eventos que son iguales o mayores que una magnitud especificada.

$P = P\,(X \geq x_T)$

Para cada ocurrencia u observación:

Éxito $\Rightarrow X \geq X_T \rightarrow$ p (Probabilidad de excedencia)

Falla $\Rightarrow X < X_T \rightarrow q = 1 - p$ (Probabilidad de no excedencia)

Periodo de retorno

La probabilidad del intervalo de recurrencia de duración T es: $(1 - p)^{T-1}p$. El valor esperado de T, tiene la siguiente expresión:

$$E(T) = \sum_{T=1}^{\infty} T(1-p)^{T-1} p$$

$$= p + 2(1-p)p + 3(1-p)^2 p + ... +$$

$$= p[1 + 2(1-p) + 3(1-p)^2 + ...]$$

Expansión en series de potencia:

$$(1+x)^n = 1 + nx + \left[\frac{n(n-1)}{2}\right]x^2 + \left[\frac{n(n-1)(n-2)}{3}\right]x^3 + ...$$

con x=-(1 - p) y n=-2

Riesgo

Reescribiendo:

$$E(T) = \frac{p}{[1-(1-p)]^2}$$

$$= \frac{1}{p}$$

$$T = \frac{1}{p}$$

Donde

p: Es la probabilidad de ocurrencia o de excedencia. $P(X \geq x_T) = \frac{1}{T_r}$

Se define el Riesgo de un diseño como la probabilidad de que la avenida para la cual se diseña la obra sea excedida.

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

La confiabilidad se define como el complemento del Riesgo:

$$\text{Confiabilidad } (\alpha) = 1\text{-}R$$