Estadística Ciencias Ambientales
Facultad de Ciencias
Departamento de Matemáticas
Profesor: Santiago de la Fuente Fernández

MUESTREO ALEATORIO

En el MUESTREO ALEATORIO estudiamos una población X con función de masa P (o función de densidad f).

El objetivo es estudiar la característica numérica X de una población.

- Si la característica X es de tipo discreto, es considerada como una variable aleatoria cuya distribución viene dada por una función de masa P.
- Si la característica X es de tipo continuo, es considerada como una variable aleatoria cuya distribución viene dada por una función de densidad f

Una muestra aleatoria de tamaño 'n' de una población X con función de masa P (o función de densidad f) es un vector aleatorio $(X_1, X_2, \cdots, X_n)$ donde:

- La distribución marginal de cada X_i viene dada por P (o por f)
- $X_1, X_2, \cdots, X_n$ son independientes

Según lo expuesto, la función de masa P (o función de densidad f) de la muestra aleatoria $(X_1, X_2, \cdots, X_n)$ viene dada por:

CASO DISCRETO: $P(x_1, x_2, \cdots, x_n) = P(x_1)\, P(x_1) \cdots P(x_n)$

CASO CONTINUO: $f(x_1, x_2, \cdots, x_n) = f(x_1)\, f(x_1) \cdots f(x_n)$

La función de masa P (o la función de densidad f) es desconocida, si fuera conocida no habría ningún problema.

El primer objetivo es obtener información sobre P (o f) a partir de las observaciones muestrales $(x_1, x_2, \cdots, x_n)$.

Para ello, tenemos que simplificar de alguna manera, los datos muestrales obtenidos. Esta necesidad nos conduce a la definición de ESTADÍSTICO.

Un Estadístico es una función real T de la muestra aleatoria $(x_1, x_2, \cdots, x_n)$.

En consecuencia, un Estadístico es una variable aleatoria $T(x_1, x_2, \cdots, x_n)$.

Algunos de los Estadísticos más utilizados son: la media muestral, la varianza muestral y la cuasi-varianza muestral.

Propiedades Estadísticos

Sea $(x_1, x_2, \cdots, x_n)$ una muestra aleatoria de una población X con esperanza μ y varianza σ^2. Entonces:

$$E(\bar{x}) = \mu \qquad E(s_x^2) = \sigma^2$$

$$E(\sigma_x^2) = \frac{n-1}{n}\sigma^2 \qquad V(\bar{x}) = \frac{\sigma^2}{n}$$

- $$E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\mu) = \mu$$

- $$V(\bar{x}) = V\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}.$$

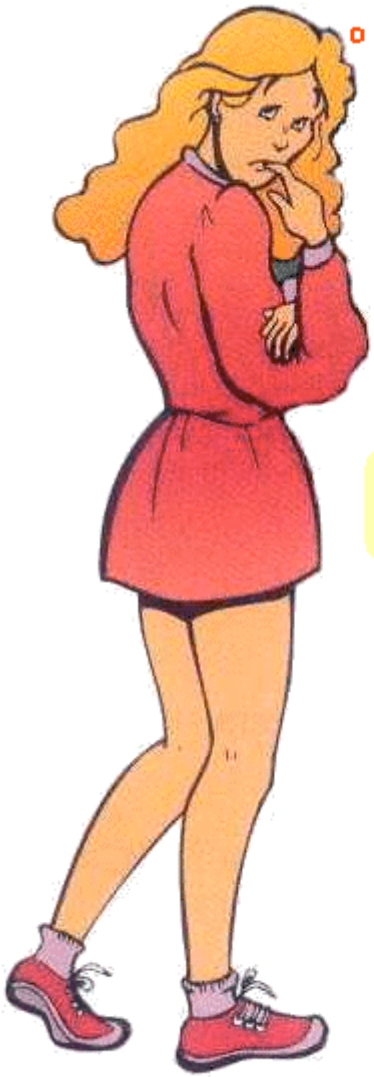
- $$E\left(s_x^2\right) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(x_i - \mu + \mu - \bar{x})^2\right] =$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left\{(x_i-\mu)^2 + (\mu-\bar{x})^2 + 2(\mu-\bar{x})(x_i-\mu)\right\}\right] =$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left((x_i-\mu)^2 + n(\bar{x}-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu)\right)\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(x_i-\mu)^2 - n(\bar{x}-\mu)^2\right] =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}E\left[(x_i-\mu)\right]^2 - n\,E\left[(\bar{x}-\mu)^2\right] = \frac{1}{n-1}\left[\left(n\sigma^2\right) - n\frac{\sigma^2}{n}\right] = \sigma^2$$

- $$E\left(\sigma_x^2\right) = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = E\left[\frac{(n-1)s_x^2}{n}\right] = \frac{(n-1)}{n}\sigma^2$$



INFERENCIA PARAMÉTRICA

En muchas ocasiones aunque la función de masa P, o de densidad f, no es conocida perfectamente, la población sometida a estudio se adapta a un determinado modelo de probabilidad (binomial, Poisson, normal, etc), donde lo único que falta por conocer es el valor de algún parámetro - p de la binomial, λ de Poisson, etc. -
En estas situaciones nos enfrentamos a un problema de Inferencia Paramétrica.

El parámetro cuyo valor desconocemos (puede ser de una o varias dimensiones) lo denotamos por θ, siendo Θ (espacio paramétrico) el conjunto de todos los posibles valores del parámetro θ.

Cuando la característica que estudiamos es de tipo discreto, la función de masa se representa por P_θ
Si es de tipo continuo, su función de densidad se representará por f_θ
La función de masa (o de densidad) de la variable aleatoria se expresa de la forma:

$$P_\theta(x_1, x_2, \cdots, x_n) = P_\theta(x_1)\, P_\theta(x_2) \cdots P_\theta(x_n)$$

$$f_\theta(x_1, x_2, \cdots, x_n) = f_\theta(x_1)\, f_\theta(x_2) \cdots f_\theta(x_n)$$

La Inferencia Paramétrica se divide en tres partes - dependiendo de la naturaleza del problema que deseemos resolver y del tipo de solución quedemos -
Estimación Puntual, Estimación por Intervalos de Confianza y Contraste de Hipótesis.

Cuando tenemos una muestra aleatoria $(X_1, X_2, \cdots, X_n)$ de una población X con función de masa P_θ (o función de densidad f_θ), es importante buscar estadísticos sencillos que conserven toda la información sobre θ que lleve la muestra $(X_1, X_2, \cdots, X_n)$. Esta cuestión origina la necesidad de definir ESTADÍSTICOS SUFICIENTES.

Sea una muestra aleatoria $(X_1, X_2, \cdots, X_n)$ de una población X con función de masa P_θ (o función de densidad f_θ).

Un estadístico T es suficiente para θ cuando:

$$P_\theta(x_1, x_2, \cdots, x_n) = g\left[T(x_1, x_2, \cdots, x_n), \theta\right] h(x_1, x_2, \cdots, x_n)$$

$$f_\theta(x_1, x_2, \cdots, x_n) = g\left[T(x_1, x_2, \cdots, x_n), \theta\right] h(x_1, x_2, \cdots, x_n)$$

Estadístico Suficiente

Una muestra aleatoria $(x_1, x_2, \cdots, x_n)$ de la población tiene como función de densidad $f_\theta(x) = \begin{cases} \theta x^{\theta-1} & \text{si } x \varepsilon (0,1) \quad \theta > 0 \\ 0 & \text{en el resto} \end{cases}$

Hallar un estadístico suficiente.

Para encontrar un estadístico suficiente T, hay que factorizar la función de verosimilitud de la forma: $L(\theta) = g(T, \theta)\, h(x_1, x_2, \cdots, x_n)$

Hacemos:

$$L(\theta) = f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n) =$$

$$= \left(\theta x_1^{\theta-1}\right)\left(\theta x_2^{\theta-1}\right) \cdots \left(\theta x_n^{\theta-1}\right) =$$

$$= \theta^n (x_1, x_2, \cdots, x_n)^{\theta-1}$$

En consecuencia,

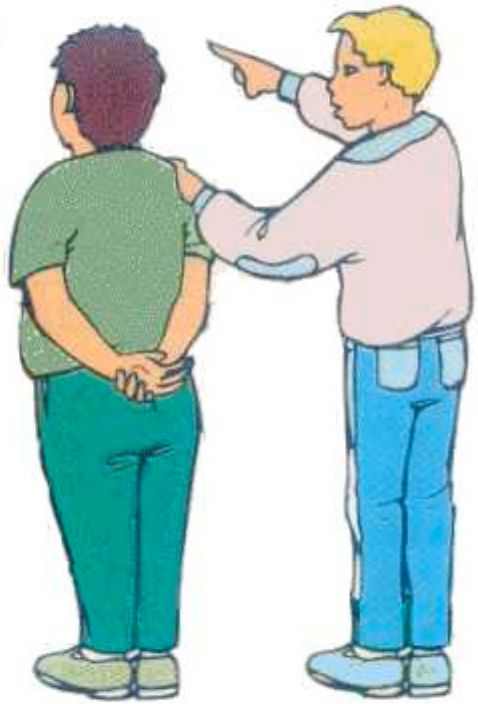
$T = (x_1, x_2, \cdots, x_n)$ es un estadístico suficiente

Estadístico Suficiente en la distribución de Poisson: $P(X=k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$

Veamos si el estadístico total muestral $\sum_{i=1}^{n} x_i$ es suficiente respecto de λ

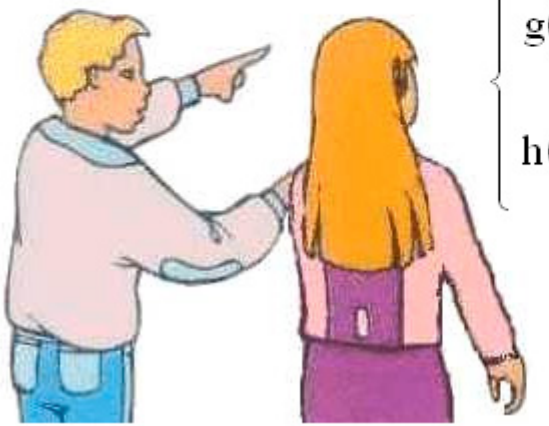
$$L(X,\lambda) = \left(\frac{\lambda^{x_1}}{x_1!} e^{-\lambda}\right)\left(\frac{\lambda^{x_2}}{x_2!} e^{-\lambda}\right) \cdots \left(\frac{\lambda^{x_n}}{x_n!} e^{-\lambda}\right) =$$

$$= \frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1!\, x_2! \cdots x_n!} e^{-n\lambda}$$



Vemos que la verosimilitud de la muestra $L(X,\lambda)$ - función de densidad conjunta de la muestra - admite una descomposición factorial, denominada también descomposición de Fisher-Neyman:

$$L(X,\theta) = g(T,\theta) \,.\, h(x_1, x_2, \cdots, x_n)$$



$$\begin{cases} g(T,\theta) = \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \\ h(x_1, x_2, \cdots, x_n) = \dfrac{1}{x_1!\, x_2! \cdots x_n!} \end{cases}$$

ESTIMACION PUNTUAL

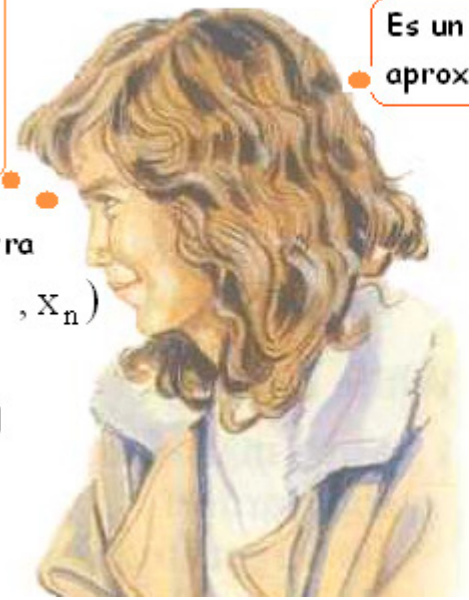
Sea una muestra aleatoria $(x_1, x_2, \cdots, x_n)$ de una población X con una distribución dada por la función de masa P_θ (o función de densidad f_θ), donde θ es un parámetro cuyo valor se desconoce - θ dentro de un conjunto Θ (espacio paramétrico) -

El objetivo en la Estimación Puntual es decidir, en función del modelo aceptado para la población y la muestra obtenida, cual es el valor de θ que nos parece más acertado.

En otras palabras,

« Se trata de elegir un punto del espacio paramétrico Θ para utilizarlo como estimación de θ »

Generalmente se estima θ $[g(\theta) = \theta]$, aunque en algunos casos se podría estimar alguna función de θ (por ejemplo, θ^2). Por este motivo se habla de estimadores puntuales.



Es un estadístico que tiene como objetivo aproximarse lo más posible al valor de $g(\theta)$

Un estimador puntual de $g(\theta)$ es una función T que a cada posible muestra $(x_1, x_2, \cdots, x_n)$ le hace corresponder una estimación $T(x_1, x_2, \cdots, x_n)$ de $g(\theta)$.

$T(x_1, x_2, \cdots, x_n)$ es una variable aleatoria (o vector aleatorio, en el supuesto de que $g(\theta)$ tenga más de una dimensión).

Se utiliza indistintamente $T \equiv T(x_1, x_2, \cdots, x_n)$

Un estimador puntual de $g(\theta)$ es una función T que a cada posible muestra $(x_1, x_2, \cdots, x_n)$ le hace corresponder una estimación $T(x_1, x_2, \cdots, x_n)$ de $g(\theta)$.

Con esa definición entran todos los estimadores. Para eliminar a los no deseados, con el ERROR CUADRÁTICO MEDIO medimos la efectividad de cada estimador.

Un ejemplo: << Entre los estimadores $T_1(x_1, x_2, \cdots, x_n) = \overline{x}$ y $T_2(x_1, x_2, \cdots, x_n) = 5$ parece razonable pensar que proporcionará estimaciones que se acercan mejor al valor del parámetro - es de esperar que la media muestral se parezca más a la esperanza θ de la población - >>

Un estimador T será más efectivo cuanto más pequeño sea su ERROR CUADRÁTICO MEDIO

$$E_\theta\left\{(T-g(\theta))^2\right\} = \begin{cases} \text{DISCRETO} \\ \displaystyle\sum \left[T(x_1, x_2, \cdots, x_n) - g(\theta)\right]^2 P_\theta(x_1)\, P_\theta(x_2) \cdots P_\theta(x_n) \\ \text{CONTINUO} \\ \displaystyle\int \left[T(x_1, x_2, \cdots, x_n) - g(\theta)\right]^2 f_\theta(x_1)\, f_\theta(x_2) \cdots f_\theta(x_n)\, dx_1\, dx_2 \cdots dx_n \end{cases}$$

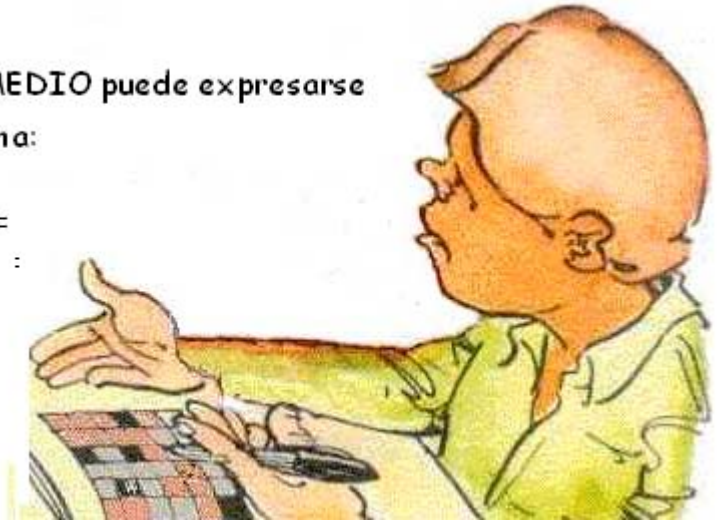
El ERROR CUADRÁTICO MEDIO puede expresarse también de la siguiente forma:

$$E_\theta\left\{(T-g(\theta))^2\right\} = E_\theta\left\{\left(T-E_\theta(T)+E_\theta(T)-g(\theta)\right)^2\right\} =$$

$$= E_\theta\left\{\left(T-E_\theta(T)\right)^2\right\} + E_\theta\left\{\left(E_\theta(T)-g(\theta)\right)^2\right\} =$$

$$= V_\theta(T) + (\text{Sesgo}(T))^2$$

$$\text{Sesgo}(T) = E_\theta(T) - g(\theta)$$



ESTIMADORES INSESGADOS

Un estimador es INSESGADO o CENTRADO cuando su valor probable coincide con el valor del parámetro a estimar. Es decir, cuando se verifica: $E_\theta(T) = g(\theta)$.
En un estimador INSESGADO el error cuadrático medio es la varianza del estimador

En caso contrario, el estimador se dice SESGADO, y a la desviación entre su valor probable y el verdadero valor del parámetro se la denomina SESGO.

Si $(x_1, x_2, \cdots, x_n)$ es una muestra aleatoria de una población X que sigue una distribución normal $N(\mu, \sigma)$, con $\theta = (\mu, \sigma)$, se tiene:

$T_1(x_1, x_2, \cdots, x_n) = \bar{x}$ es un estimador insesgado de $g_1(\theta) = \mu$, puesto que $E_\theta(\bar{x}) = \mu$

$T_2(x_1, x_2, \cdots, x_n) = s_x^2$ es un estimador insesgado de $g_2(\theta) = \sigma^2$, puesto que $E_\theta(s_x^2) = \sigma^2$

¿ IMPORTA EL TAMAÑO DE LA MUESTRA ?

La lógica nos induce a pensar que cuanto más grande sea el número de observaciones, más próxima estará la estimación $T(x_1, x_2, \cdots, x_n)$ del verdadero valor de $g(\theta)$

Un estimador es consistente para estimar $g(\theta)$ cuando verifica las propiedades:

$$\lim_{n \to \infty} E_\theta(T) = g(\theta) \quad \forall\, \theta \in \Theta \qquad \lim_{n \to \infty} V_\theta(T) = 0 \quad \forall\, \theta \in \Theta$$

MÉTODO DE LOS MOMENTOS ....

Si $(x_1, x_2, \cdots, x_n)$ es una muestra aleatoria de una población X con función de masa P_θ (o función de densidad f_θ), siendo el parámetro $\theta = (\theta_1, \theta_2, \cdots, \theta_n)$.

El estimador esta formado por los valores $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n)$ que se obtienen al resolver en $\theta_1, \theta_2, \cdots, \theta_n$ el sistema de ecuaciones:

$$\left[\begin{array}{c} E_\theta(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i \\ \cdots\cdots\cdots\cdots \\ E_\theta(x^k) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^k \end{array}\right]$$

El método presenta inconvenientes - puede ocurrir que la estimación obtenida corresponda a valores que están fuera del espacio paramétrico.

El método más utilizado es el de máxima verosimilitud.

MÉTODO DE MÁXIMA VEROSIMILITUD

Si $(x_1, x_2, \cdots, x_n)$ es una muestra aleatoria de una población X con función de masa P_θ (o función de densidad f_θ), siendo el parámetro $\theta = (\theta_1, \theta_2, \cdots, \theta_n)$

El estimador esta formado por los valores $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n)$ que maximizan lo que llamamos función de verosimilitud de la muestra $(x_1, x_2, \cdots, x_n)$ obtenida:

$$L(\theta) = L(\theta; x_1, x_2, \cdots, x_n) = \begin{cases} P_\theta(x_1)P_\theta(x_2) \cdots, P_\theta(x_n) & \text{DISCRETO} \\ f_\theta(x_1)f_\theta(x_2) \cdots, f_\theta(x_n) & \text{CONTINUO} \end{cases}$$

La estimación de máxima verosimilitud siempre es un valor del espacio paramétrico.

La función de verosimilitud expresa la probabilidad (o densidad) que los diferentes valores de θ dan a la muestra obtenida - En consecuencia, maximizamos esa probabilidad (o densidad).

Muchas veces la forma más sencilla de encontrar el estimador de máxima verosimilitud es considerar la función $[\log L(\theta)]$ - el logaritmo de $L(\theta)$, ya que se maneja más fácilmente y presenta los mismos máximos y mínimos -.

En el sistema de ecuaciones adjuntos se despejan los valores $(\theta_1, \theta_2, \cdots, \theta_n)$

$$\begin{cases} \dfrac{\vartheta \log L(\theta)}{\vartheta \theta_1} = 0 \\ \cdots\cdots \\ \cdots\cdots\cdots \\ \dfrac{\vartheta \log L(\theta)}{\vartheta \theta_k} = 0 \end{cases}$$