TEMA 3: ESTIMACIÓN PUNTUAL.

3.1.- FUNDAMENTOS.

La **inferencia estadística** proporciona un método objetivo que establece reglas base para criticar, rechazar y aceptar "items" de información científica cuando prevalecen condiciones de incertidumbre. De esta forma, se pueden extraer conclusiones (con incertidumbre valorada mediante probabilidad si se aplican principios y reglas objetivables) sobre una población utilizando como ñmateria primaò la información muestral. Una inferencia es una extensión de lo particular a lo general. La inferencia inductiva es un proceso con riesgo ya que una inferencia inductiva exacta es imposible. Una clara aplicación se tiene sobre la realidad económica de la que se estudian problemas concretos para mejorar el conocimiento de ciertas poblaciones (grandes grupos de individuos, miles o millones) mediante encuestas que recogen información de una muestra (pequeños grupos de individuos extraídos de la población, cientos).

Los procedimientos empleados pueden clasificarse en función del objetivo de la inferencia en métodos paramétricos (se desea evaluar los parámetros poblacionales desconocidos) y métodos no paramétricos (se desea conocer otra característica de la población). Los procedimientos empleados pueden clasificarse según la técnica inferencial en Estimación y Contraste de Hipótesis. La Estimación asigna valores a parámetros: si asigna un único valor es Estimación Puntual y si asigna un intervalo es Estimación por Intervalos. El Contraste de Hipótesis establece una regla de rechazo o aceptación de cierta afirmación o hipótesis, sobre los parámetros poblacionales o sobre otro aspecto poblacional no paramétrico, mediante prueba basada en información muestral. Los procedimientos empleados pueden clasificarse según el tipo de información que utilizan en métodos clásicos (los parámetros se consideran valores fijos pero desconocidos) y métodos bayesianos (los parámetros se tratan como si fueran variables aleatorias).

Se pretende asignar un valor a un **parámetro** poblacional. El parámetro es un valor representativo de una población (visión generalista o descriptiva). Desde una visión más formal o inferencial, es un valor que define una familia de objetos matemáticos o modelos de probabilidad. Esto es, es un valor asociado a la población, entendida como variante, que permite fijar el modelo estocástico. Hay modelos con un parámetro (uniparamétrico) como $B(1,p)$, χ^2_n, t_n o $Poisson\ \lambda$; modelos con dos (biparamétricos) como $B(n,p)$, $U(a,b)$, $N(\mu,\sigma)$ o $F_{m,n}$; y modelos con más de dos parámetros.

Para conocer el parámetro se utilizan los **estimadores**. El Estimador será un estadístico (cualquier función muestral) dedicado al conocimiento de un parámetro poblacional desconocido (útil para estimar). El parámetro es una característica de la población, la información sobre la población se “transmite” a la muestra mediante el proceso de muestreo, la muestra es resumida (manteniendo la información “relevante” o importante) por el estadístico, el estadístico seleccionado por sus cualidades se emplea para estimar y se denomina estimador. De esta forma, parámetro y estimador están “conectados” o relacionados.

Además, se tiene que el estimador será una variable aleatoria antes de obtener la muestra concreta (a priori) y será un valor o concreción de la variable aleatoria después de obtener la muestra (a posteriori). En la gran mayoría de las ocasiones los parámetros poblacionales a estimar son, como es lógico, los más importantes: la media y la varianza. Por ello, parece de sentido común que la media muestral, la varianza muestral y la cuasivarianza muestral sean estadísticos relevantes (por analogía) o estadísticos estimadores cuya distribución de probabilidad en el muestreo necesita ser conocida por su amplia utilización (por eso se estudia en el tema anterior). Así se pueden plantear las siguientes estimaciones puntuales:

Parámetro poblacional	**Estimador**	**Estimación**

Para la elección de estos estimadores puntuales la base ha sido, principalmente la intuición y la posible analogía de los parámetros poblacionales con sus correspondientes valores muestrales, pero éste no será el método más adecuado para la obtención de estimadores puntuales, aunque en este caso se obtienen estimadores satisfactorios para los parámetros poblacionales. En general, el problema de obtener estimadores puntuales no será tan sencillo, por ello se emplean las propiedades que serían deseables que se cumplieran por los diferentes estimadores puntuales obtenidos, aunque no existe un mecanismo o método único que nos permita obtener el mejor estimador puntual en todas las circunstancias.

Resumiendo, el **proceso de estimación puntual** sería:

1.- Se tiene una población que presenta parámetro o parámetros desconocidos (por ejemplo una población normal de media desconocida y varianza 4).

2.- Se plantea asignar un único valor al parámetro o parámetros desconocidos (por ejemplo interesa dar un único valor a la media poblacional).

3.- Se planifica un muestreo probabilístico, en este curso siempre m.a.s., para obtener una muestra $X(X_1,....,X_n)$ (por ejemplo m.a.s. de tamaño ñ4ò).

4.- Se selecciona con algún criterio un estadístico estimador $\hat{\theta}(X_1,....,X_n)$ que a priori es una variable aleatoria (por ejemplo la variante media muestral).

5.- Se obtiene la muestra concreta $x(x_1,....,x_n)$ tras efectuar el muestreo (por ejemplo x_1=1, x_2=1, x_3=2, x_4=2).

6.- Se sustituye en el estadístico estimador la muestra obtenida $\hat{\theta}(x_1,....,x_n)$ que a posteriori es una concreción o suceso denominado estimación puntual (por ejemplo la $\bar{x} = \dfrac{1+1+2+2}{4} = 1{,}5$).

7.- Esa estimación puntual es el valor concreto que se asigna al parámetro desconocido del verdadero valor del parámetro poblacional (según el enfoque de la inferencia clásica).

Para conseguir, al menos, una técnica objetiva y ñbuenaò de estimación puntual se debe garantizar que los estadísticos estimadores cumplen propiedades deseables. Por eso se estudian las propiedades adecuadas para los estimadores: insesgo (el estimador en media es el parámetro a estimar o esperanza del estimador coincide con el valor del parámetro desconocido), eficiencia (el estimador tiene poca variabilidad o minimización de la varianza del estimador), consistencia (estimaciones mejores cuanto mayor sea el número de unidades observadas o convergencia en probabilidad), suficiencia (el estadístico contiene toda la información muestral relevante respecto del parámetro desconocido), robustez (cambios en las hipótesis de comportamiento poblacional no afectan al estimador), etc.

Ahora bien, ¿cómo seleccionar los estimadores más adecuados para luego evaluar sus propiedades?. Cualquier función de la muestra es un posible estimador y son infinitas las posibilidades. El problema es mayor cuando existen muchos estimadores que puedan ser ñobjeto de estudioò. El criterio de ñanalogíaò tiene una débil justificación formal. Deben emplearse métodos para construir a priori buenos estimadores (sólo el mejor o unos pocos mejores), aunque a posteriori se compruebe si cumplen las propiedades deseables.

Conclusión: se necesita disponer de métodos objetivos para construir estimadores que permitan obtener, de entre los infinitos estimadores, los más razonables. Los métodos propuestos suministrarán unos estimadores que cumplirán, al menos, ciertas deseables propiedades. Existen varios métodos de estimación, entre ellos: momentos y máxima verosimilitud.

3.2.- MÉTODO DE LOS MOMENTOS.

y aplicable a muchas situaciones. A veces puede llevar a resultados incompatibles con la población estudiada dado que no utiliza toda la información del modelo poblacional.

Método. El método consiste en establecer una igualdad entre momentos muestrales y poblacionales (respecto del origen). Los momentos poblacionales, en general, estarán en función de los parámetros desconocidos. Se plantean tantas ecuaciones como incógnitas (número de parámetros poblacionales desconocidos) con el fin de llegar a un sistema de ecuaciones completo compatible que permita obtener la estimación de los parámetros como resultado del sistema de ecuaciones.

$$\left.\begin{array}{l} \alpha_1(\theta_1,\ldots,\theta_k) = a_1 \\ \alpha_2(\theta_1,\ldots,\theta_k) = a_2 \\ \ldots \\ \alpha_k(\theta_1,\ldots,\theta_k) = a_k \end{array}\right\} \Rightarrow (\hat{\theta}_1,\ldots,\hat{\theta}_k)$$

Justificación. Este procedimiento de estimación se justifica, siguiendo el teorema de Kintchine, por la convergencia en probabilidad de los momentos muestrales a los poblacionales (supuesta la i.i.d. de las variantes que se tienen en los elementos de la muestra) y en que los momentos muestrales respecto del origen son estimadores insesgados de los momentos poblacionales respecto del origen.

Propiedades. En cuanto a las propiedades del estadístico estimador es claro que si se estima un momento respecto del origen de la distribución poblacional se llega a que el estimador es insesgado. Por el mencionado teorema de Kintchine, si se estima un momento respecto del origen de la distribución poblacional se tienen estimaciones consistentes y con el teorema de Lindeberg-Levi se comprueba la normalidad asintótica.

Si se considera en general (no necesariamente estimando un momento poblacional), los estimadores son consistentes pero no son centrados (insesgados) ni de varianza mínima.

La mayor ventaja es su simplicidad. La mayor desventaja es que no considera toda la información disponible sobre la distribución poblacional y sobre la muestra (sólo

EJEMPLO.

Hallar, por el método de los momentos, el estimador del parámetro θ de una población, supuesto que se extrae una m.a.s. de tamaño “n”, sabiendo que la función de densidad poblacional es: $f(x) = f(x,\theta) = \dfrac{\theta}{(1+x)^{1+\theta}} \quad x \geq 0,\ \theta > 0.$

Como sólo hay un parámetro desconocido se plantea una ecuación con una incógnita: primer momento poblacional con respecto al origen (que es función de $\boldsymbol{\theta}$*) se iguala a primer momento muestral con respecto al origen).*

$$\left.\begin{array}{l} \alpha_1(\theta) = \displaystyle\int_{-\infty}^{+\infty} x \frac{\theta}{(1+x)^{1+\theta}}\, dx = \frac{1}{\theta - 1} \\ a_1 = \displaystyle\frac{1}{n}\sum_{i=1}^{n} x_i \end{array}\right\} \Rightarrow \frac{1}{\theta-1} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Al despejar se obtiene el estimador $\hat{\theta} = \dfrac{n}{\sum_{i=1}^{n} x_i} + 1.$

3.3.- MÉTODO DE MÁXIMA VEROSIMILITUD.

Verosimilitud. Conviene introducir el concepto de verosimilitud y la función de verosimilitud como instrumento básico para inferencias en grandes muestras.

La verosimilitud indicará la confianza que se tiene en la estimación o inferencia sobre un parámetro atendiendo a la información contenida en la muestra (según la función de verosimilitud) y considerando que habitualmente sucede lo más probable. Con base en el conocimiento que se tiene (o se supone válido) de la distribución de probabilidad poblacional se construye la función de verosimilitud para expresar mediante ésta la posibilidad u orden de preferencia en cuanto al valor concreto que pueda tomar el parámetro desconocido.

En el caso discreto, se obtiene la función de verosimilitud como probabilidad de

$$L(\theta; X = x) = L(\theta / X = x) = kP(\theta / X = x).$$

La probabilidad de la muestra es

$$P(X = x) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n) = \prod_{j=1}^{n} P(X_j = x_j) = \prod_{j=1}^{n} P(\xi = x_j)$$

y la función de verosimilitud (eliminando la constante del efecto escala o unidades muestrales) es

$$L(x_1,\ldots,x_n, \boldsymbol{d}) = P(\xi = x_1, \boldsymbol{d}) \cdot\ P(\xi = x_2, \boldsymbol{d}) \cdot \ldots\ \cdot P(\xi = x_n, \boldsymbol{d})$$

$$L(x_1,\ldots,x_n, \boldsymbol{d}) = \prod_{i=1}^{n} P(\xi = x_i, \theta)$$

En el caso continuo, se obtiene la función de verosimilitud como densidad de la muestra pero ñentendidaò como función del parámetro (la variable de la función es el parámetro). Dada una m.a.s. *x(x₁,x₂,é ,xₙ) y* la función de densidad poblacional $f_\xi(x,\theta)$ se llega a una función de verosimilitud definida como proporcional a la densidad de aparición de la muestra condicionada al parámetro

$$L(\theta; X = x) = L(\theta / X = x) = k \cdot f(\theta / X = x).$$

La densidad de la muestra es

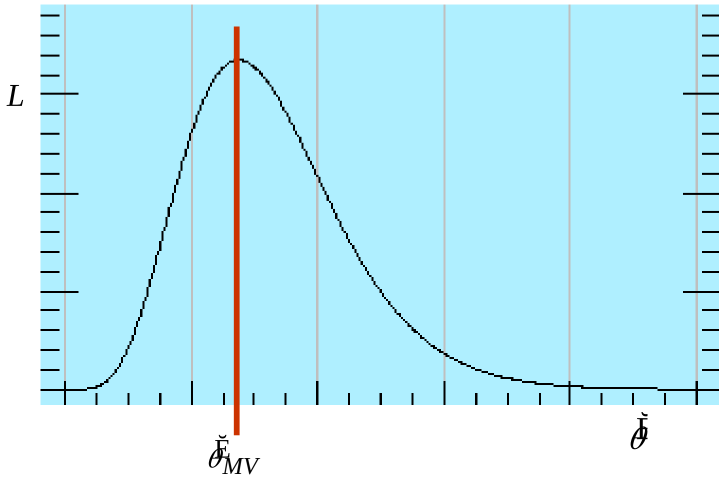
$$f_X(x) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \ldots \cdot f_{X_n}(x_n) = \prod_{j=1}^{n} f_{X_j}(x_j) = \prod_{j=1}^{n} f_\xi(x_j)$$

y la función de verosimilitud (eliminando la constante del efecto escala o unidades muestrales) es

$$L(x_1,\ldots,x_n, \boldsymbol{d}) = f_\xi(x_1, \boldsymbol{d}) \cdot\ f_\xi(x_2, \boldsymbol{d}) \cdot \ldots\ \cdot f_\xi(x_n, \boldsymbol{d})$$

$$L(x_1,\ldots,x_n, \boldsymbol{d}) = \prod_{i=1}^{n} f_\xi(x_i, \theta)$$

En el caso discreto, la probabilidad muestral (probabilidad de una muestra concreta) se obtiene en muestreo aleatorio simple como producto de las probabilidades de cada elemento muestral (probabilidad de un elemento muestral concreto). En el caso continuo se utiliza los valores referidos a la función de densidad.



Ante una muestra concreta, permite elegir o indicar preferencia entre dos parámetros: $\theta_1 \succ \theta_2 \quad si \quad L(\theta_1 / X) > L(\theta_2 / X)$.

De hecho, toda la información relevante de la muestra que permite elegir entre dos parámetros se encuentra en la razón de verosimilitud o cociente de las verosimilitudes (se elimina el efecto escala o unidades de los elementos muestrales que aparece al incorporar los valores muestrales obtenidos):

$$\lambda(X) = \frac{L(\theta_1 / X)}{L(\theta_2 / X)}.$$

Si $\lambda(X) > 1$ se prefiere el parámetro del numerador, si $\lambda(X) < 1$ se prefiere el parámetro del denominador, y si $\lambda(X) = 1$ hay indiferencia (no preferencia).

Con grandes muestras, tal como se tienen en muchos casos empíricos reales, la verosimilitud tiene intuitivamente mucha importancia.

Idea intuitiva. Habitualmente sucede lo más probable, con lo que la muestra concreta se ñha obtenidoò porque presentaba la máxima probabilidad. De forma más intuitiva se puede decir que ñhabitualmente sucede lo más probableò y se tienen o dan los ñvalores muestrales más probablesò (cuanto mayores sean los tamaños muestrales mejores serán las estimaciones ya que ñpesaò más la muestra y más ñadecuadaò será esta idea intuitiva).

poblacional se construye la función de verosimilitud para expresar mediante ésta la posibilidad u orden de preferencia en cuanto al valor concreto que pueda tomar el parámetro desconocido.

Método. El método consiste en obtener el valor del parámetro (dentro del espacio paramétrico) que maximiza la función de verosimilitud. Para facilitar los cálculos se recurre al logaritmo neperiano de la función de verosimilitud (denominado función soporte) ya que, aunque proporciona un máximo distinto, éste se encontrará en el mismo punto (o valor estimador del parámetro):

$$\ln L(\hat{\theta}; X) = \underset{\theta \in \Theta}{máx} \ln L(\theta; X).$$

En la práctica es más fácil maximizar el logaritmo de la función de verosimilitud (maximizar la denominada función soporte).

En resumen: se parte de una muestra (m.a.s.) y de una población con función de probabilidad conocida, se construye la función de verosimilitud, se toman logaritmo y se halla la función soporte, se busca el valor del parámetro que hace máximo el valor de la función (si diferenciable condiciones de primer y segundo orden). Ese valor del parámetro es el estimador si se deja en función de los valores muestrales y es la estimación puntual si se sustituyen los valores concretos de la muestra dada.

Justificación. Este procedimiento de estimación se justifica por el principio de verosimilitud. El principio de verosimilitud mantiene que las inferencias sobre el parámetro efectuadas con la muestra son las mismas que las realizadas con la función de verosimilitud y que el cálculo de esta función de verosimilitud supone una reducción del espacio muestral.

Propiedades. El método de máxima verosimilitud tiene gran importancia debido a las adecuadas propiedades asintóticas que presenta. Esto es, con grandes muestras, tal como se tienen en muchos casos empíricos reales, los estimadores de máxima verosimilitud tienen interesantes propiedades. Son asintóticamente insesgados y

EJEMPLO.

Hallar, por el método de máxima verosimilitud, el estimador del parámetro ď de una población, supuesto que se extrae una m.a.s. de tamaño ñnò, sabiendo que la función de densidad poblacional es: $f(x) = f(x,\theta) = \dfrac{\theta}{(1+x)^{1+\theta}} \qquad x \geq 0,\ \theta > 0.$

Para obtener el estimador se comprueba que en el enunciado se da m.a.s. y función de probabilidad poblacional conocida (en este caso función de densidad por ser variante poblacional continua).

Se construye la función de verosimilitud y se toman logaritmo para hallar la función soporte

$$L(x_1,\ldots,x_n,\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \frac{\theta^n}{\prod_{i=1}^{n}(1+x_i)^{1+\theta}}$$

$$Ln\,L(x_1,\ldots,x_n,\theta) = nLn\theta - (1+\theta)\sum_{i=1}^{n} Ln(1+x_i)$$

Se busca el valor del parámetro que hace máximo el valor de la función (primera derivada respecto al parámetro igual a cero y segunda derivada respecto al parámetro ñevaluadaò menor que cero)

$$\frac{\partial Ln\,L(x_1,\ldots,x_n,\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} Ln(1+x_i) = 0$$

$$\Rightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^{n} Ln(1+x_i)}$$

$$\left.\frac{\partial^2 Ln\,L(x_1,\ldots,x_n,\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}^2} < 0$$

Ese valor del parámetro es el estimador ya que se deja en función de los valores muestrales. Si tuviéramos muestra concreta dada se llega a la estimación puntual sustituyendo.

probabilístico de éstos pero no permiten efectuar ninguna afirmación sobre las estimaciones particulares). Esto es debido a la aleatoriedad del muestreo y al carácter siempre desconocido del verdadero valor del parámetro poblacional (según el enfoque de la inferencia clásica). En el tema de estimación por intervalos se propondrá una solución a este problema: cada estimación puntual se debe acompañar de un intervalo de confianza.