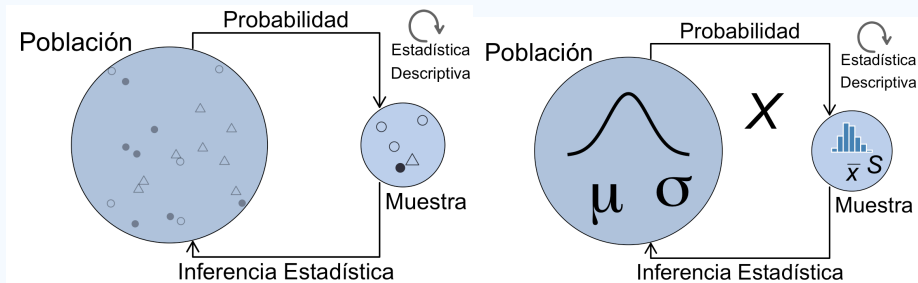


Inferencia Estadística. Estimación puntual

Inferencia Estadística - Grado en Ciencia e Ingeniería de Datos

Curso académico 2024-2025





- X es una variable aleatoria (¡de interés!) cuya función de distribución es conocida a excepción de determinados parámetros (Por ejemplo, μ y σ en la distribución Normal)
- “Conocidos” (estimados) los parámetros (por ejemplo, \bar{x} y S para la Normal) \rightarrow Función de distribución de probabilidad totalmente determinada y ¡a extraer información y sacar conclusiones!

- Sabemos que el número de horas al día que pasan las personas en Instagram sigue una distribución Normal. Estamos interesados en conocer la media de horas.
- Variable aleatoria X : horas que pasan al día los estudiantes de la URJC en Instagram
- Población: Estudiantes de la URJC
- Para estimar la media μ , se pregunta a 10 estudiantes al azar el número de horas que pasan al día x_1, \dots, x_{10} :

2.5, 3.0, 1.5, 4.0, 2.0, 3.5, 2.8, 1.0, 4.5, 3.2

- X variable aleatoria (v.a.) con función de distribución F_θ conocida, que depende de un parámetro $\theta \in \Theta \subseteq \mathbb{R}$ desconocido, siendo Θ su espacio paramétrico
- El objetivo de la inferencia estadística es lograr buenas estimaciones/aproximaciones del parámetro θ en base a una muestra aleatoria simple (m.a.s.) X_1, \dots, X_n de la población. $X_i, i = 1, \dots, n$ son v.a. independientes e igualmente distribuidas F_θ
- Con cada observación x_1, \dots, x_n de la muestra aleatoria simple se puede hacer una estimación del parámetro desconocido $\theta \rightarrow$ estadístico muestral o estimador

- Formalmente, dada una variable aleatoria X con función de distribución F_θ , con θ desconocido, un estadístico o estimador $T = T(X_1, \dots, X_n)$ es una función real de la m.a.s. (X_1, \dots, X_n) que estima el valor del parámetro desconocido

$$T = T(X_1, \dots, X_n) = \hat{\theta}$$

- Ejemplos:
 - Media, mediana
 - Varianza, desviación estándar
- Un estadístico es una variable aleatoria, y por lo tanto, tiene asociada una distribución que se denomina **distribución muestral**

Dado el mismo estadístico, la aplicación del mismo a distintas muestras concretas dará lugar a diferentes **estimaciones**

Ejemplo Instagram:

- Estadístico media muestral \bar{X}
- Muestra A: 2.5, 3.0, 1.5, 4.0, 2.0, 3.5, 2.8, 1.0, 4.5, 3.2 $\rightarrow \bar{x}_A = 2.8$
- Muestra B: 3.1, 2.7, 1.8, 4.2, 3.3, 2.4, 1.6, 4.0, 3.5, 2.9 $\rightarrow \bar{x}_B = 2.95$
- Muestra C: 2.9, 3.6, 1.4, 4.1, 2.3, 3.0, 1.9, 4.4, 3.8, 2.2 $\rightarrow \bar{x}_C = 2.96$

La media muestral varía en las distintas muestras \rightarrow es una variable aleatoria con una distribución de probabilidad. Caracterizaremos dichas distribuciones muestrales mediante su media y su varianza.

- Media: $T_1(X_1, \dots, X_n) = \overline{X} = \frac{(X_1 + \dots + X_n)}{n}$
- Mediana: $T_2(X_1, \dots, X_n) = \frac{X_{(n/2)} + X_{(n/2+1)}}{2}$
- Media del máximo y el mínimo:
 $T_3(X_1, \dots, X_n) = \frac{\max(X_1, \dots, X_n) + \min(X_1, \dots, X_n)}{2}$
- etc

Distintos estadísticos pueden estimar el valor del mismo parámetro

No todos los estadísticos sirven para estimar el mismo parámetro

La aplicación de cada estimador a la muestra ofrece una estimación diferente de la media poblacional μ :

- Media:

$$T_1(x_1, \dots, x_{10}) = \bar{x} = \frac{(2.5+3.0+1.5+4.0+2.0+3.5+2.8+1.0+4.5+3.2)}{10} = 2.8$$

- Mediana: $T_2(x_1, \dots, x_{10}) = \frac{2.8+3.0}{2} = 2.9$

- Media del máximo y el mínimo: $T_3(x_1, \dots, x_{10}) = \frac{4.5+1}{2} = 2.75$

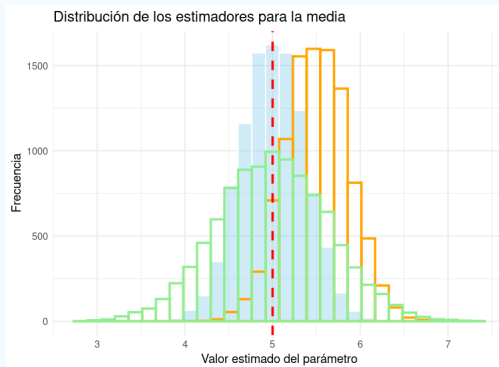
Así, se obtienen 3 estimaciones de la media poblacional: $\hat{\mu}_1 = 2.8$,
 $\hat{\mu}_2 = 2.9$, $\hat{\mu}_3 = 2.75$

- En el tema 1 dijimos: $DATOS = MODELO + ERROR$
- Particularmente: $ERROR = |T(X_1, \dots, X_n) - \theta|$
- Queremos que el error sea lo mínimo posible, pero desconocemos θ
→ Pedimos ciertas propiedades a los estimadores para que sean adecuados para estimar un parámetro

Supongamos una población Normal con media conocida $\mu = 5$

Se obtiene una muestra de 10000 observaciones y se calculan los 3 estadísticos que se muestran a continuación (azul, naranja y verde).

¿Qué opinas?



- **Insesgadez.** Un estimador es insesgado (o centrado) si, en promedio, coincide con el valor verdadero del parámetro que se estima. Es decir, el valor esperado del estimador es igual al parámetro poblacional:
 $E[\hat{\theta} = T(X_1, \dots, X_n)] = \theta$. Esto quiere decir que la distribución del estadístico muestral está centrada en el verdadero valor del parámetro
- **Eficiencia.** La varianza de un estimador debe ser lo más pequeña posible. Entre dos estimadores insesgados, el más eficiente es el que tiene menor varianza, es decir, el que proporciona estimaciones más precisas.

- **Consistencia:** Un estimador es consistente si, a medida que el tamaño de la muestra aumenta, la estimación se aproxima al valor verdadero del parámetro. Es decir:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \geq \delta) = 0, \forall \delta > 0$$

donde n es el tamaño muestral

- **Suficiencia:** Un estimador del parámetro θ es suficiente si utiliza toda la información contenida en la muestra sobre el parámetro que se está estimando. Es decir, si la distribución de la muestra X_1, \dots, X_n dado el estadístico T es independiente del valor del parámetro desconocido. Esto es, se tiene la misma información sabiendo el valor del estadístico T que conociendo todas las observaciones de la muestra

El Teorema Central del Límite (TCL) establece que, bajo ciertas condiciones, la distribución de la suma (o el promedio) de un número (lo suficientemente grande) de variables aleatorias independientes e idénticamente distribuidas tiende a seguir una distribución Normal, independientemente de la distribución original de las variables.

Formalmente, el TCL establece que si, X_1, X_2, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas, con media μ y varianza $\sigma^2 < \infty$, entonces para n suficientemente grande se verifica

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Usamos estadísticos para estimar los parámetros desconocidos
- Queremos que estos estadísticos tengan buenas propiedades
- Estos estadísticos son variables aleatorias con distribución de probabilidad

--> Vamos a estudiar algunos de los estadísticos más comunes, sus distribuciones y características

¡Ojo! Por el TCL, teniendo el tamaño muestral adecuado, sabremos que la distribución de los estadísticos será una Normal

Dadas X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 conocida,

la media muestral

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

tiene las siguiente esperanza (media) y varianza:

- $E[\bar{X}] = \mu$
- $V[\bar{X}] = \frac{\sigma^2}{n}$

Entonces, con n suficientemente grande, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

La varianza muestral, que sirve para estimar la varianza poblacional, es:

$$V^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

es una variable aleatoria con media:

$E[V^2] = \frac{n-1}{n} \sigma^2 \rightarrow$ ¡No es un estimador insesgado de la varianza poblacional!

¿Ideas?

La cuasivarianza muestral, que sirve para estimar la varianza poblacional, es:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

es una variable aleatoria con media:

$E[S^2] = \sigma^2 \rightarrow$ Estimador insesgado de la varianza poblacional

$$V[S^2] = \frac{2\sigma^4}{n-1}$$

La **estimación puntual** es una técnica en estadística que consiste en utilizar los datos de una muestra para calcular un valor único, denominado **estimador puntual**, que se usa como mejor aproximación de un parámetro desconocido de la población. Nótese que la función de distribución de la población es conocida a excepción de dicho parámetro.

La estimación puntual proporciona una forma simple y directa de hacer inferencias sobre parámetros poblacionales a partir de una muestra, aunque su simplicidad también implica que no proporciona información sobre la precisión o variabilidad de la estimación, aspectos que se abordan mediante la **estimación por intervalos** y otras técnicas inferenciales.

- Media muestral \bar{X} es un estimador insesgado para la media poblacional μ , cuya distribución cumple que:
 - $E[\bar{X}] = \mu$
 - $V[\bar{X}] = \sigma^2/n$
- Dos casos:
 - Varianza conocida y n suficientemente grande ($n \geq 30$):
 $\bar{X} \sim N(\mu, \sigma/\sqrt{n}) \Leftrightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
 - Varianza desconocida y n suficientemente grande (se requiere normalidad), se utiliza V^2 y entonces $\frac{\bar{X}-\mu}{V/\sqrt{n}} \sim t_{n-1}$ siendo t_{n-1} la distribución t-Student con $n - 1$ grados de libertad

- Supongamos que se quiere estimar el gasto promedio que hacen los estudiantes en cultura
- Se selecciona una muestra aleatoria de 5 estudiantes cuyos gastos mensuales (en euros) son: 100, 50, 125, 20 y 40
- ¿Estimación puntual de la media?

- Se usa la cuasivarianza porque es insesgado
- Cuasivarianza muestral $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- Su distribución cumple:
 - $E[S^2] = \sigma^2$
 - $V[S^2] = \frac{2\sigma^4}{n-1}$
- En poblaciones normales o con n suficientemente grande cumple

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

siendo χ_{n-1}^2 la distribución chi-cuadrado con $n-1$ grados de libertad

- Supongamos que se quiere estimar el gasto promedio que hacen los estudiantes en cultura
- Se selecciona una muestra aleatoria de 5 estudiantes cuyos gastos mensuales (en euros) son: 100, 50, 125, 20 y 40
- ¿Estimación puntual de la varianza?
- Supongamos que la muestra es de tamaño 500 y tenemos la media y la varianza estimadas. ¿Qué más sabemos?

Dadas X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con una distribución Bernoulli de parámetro p . En este caso $\hat{p} = \bar{X}$ es un estimador insesgado para p y su distribución muestral cumple:

- $E[\hat{p}] = p$
- $V[\hat{p}] = \frac{p(1-p)}{n}$

Si $n \geq 30$, por el TCL se tiene que

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \Leftrightarrow \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0, 1)$$

- En una encuesta a 100 personas, 70 dicen que prefieren estudiar de noche y 30 de día
- Se quiere estimar la proporción de personas que prefieren estudiar de noche
- ¿Estimación puntual de la proporción?

- Método de los momentos
- Método de la máxima verosimilitud

El método de los momentos es una técnica utilizada en estadística para estimar los parámetros desconocidos $(\theta_1, \dots, \theta_k)$ de una distribución de probabilidad. Fue introducido por el estadístico *Karl Pearson* en 1984.

Este método se basa en igualar los momentos muestrales (calculados a partir de los datos observados) con los momentos teóricos (expresados en términos de los parámetros de la distribución poblacional) y, con el sistema de ecuaciones resultantes, despejar los parámetros desconocidos. Así se logran los estimadores de los parámetros poblacionales desconocidos.

En estadística, los **momentos poblaciones** de una distribución son medidas que describen diversas características de la misma, como su media, varianza, simetría y curtosis. Los momentos más comunes son:

- ➊ Primer momento (media): $\mu = E[X]$
- ➋ Segundo momento (varianza): $\mu_2 = E[X^2]$
- ➌ Tercer momento (asimetría): $\mu_3 = E[X^3]$
- ➍ Cuarto momento (curtosis): $\mu_4 = E[X^4]$
- ➎ k -ésimo momento: $\mu_k = E[X^k]$

Recordemos que:

- $\mu_k = E[X^k] = \sum_{i=1}^n x_i^k p_i$, siendo p_i la probabilidad de cada valor del dominio de la v.a. en el caso discreto (función de masa)
- $\mu_k = E[X^k] = \int x^k f(x) dx$ siendo $f(x)$ la función de densidad en el caso continuo

Los momentos muestrales se definen como:

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

- ➊ **Calcular Momentos Muestrales:** Se calculan los momentos muestrales de los datos observados. El (k) -ésimo momento muestral se define como $m_k = \frac{1}{n} \sum_{i=1}^n x_i^k$ donde n es el tamaño de la muestra y x_i son los valores de la muestra
- ➋ **Igualar Momentos Muestrales y Teóricos:** Se igualan los momentos muestrales con los momentos teóricos de la distribución. Los momentos teóricos se expresan en términos de los parámetros desconocidos que se desean estimar
- ➌ **Resolver el Sistema de Ecuaciones:** Se resuelve el sistema de ecuaciones resultante para encontrar los estimadores de los parámetros desconocidos. Fíjate que tenemos k ecuaciones y k parámetros $(\theta_1, \dots, \theta_k)$. De modo que es posible despejar los parámetros de estas ecuaciones, quedando estos parámetros en función de los momentos. En estas ecuaciones se sustituyen los momentos poblacionales por sus correspondientes momentos muestrales. Esto da como resultado estimaciones de esos parámetros.

¡Veámos cómo obtener estimadores para los parámetros μ y σ^2 de la distribución Normal con el método de los momentos!

Ventajas

- **Simplicidad:** El método de los momentos es relativamente sencillo de aplicar y no requiere técnicas complejas de optimización.
- **Intuición:** Ofrece una interpretación intuitiva de los parámetros en términos de momentos.

Limitaciones

- **Precisión:** Los estimadores de los momentos no siempre son los estimadores más eficientes (no tienen la mínima varianza posible).
- **Aplicabilidad:** En algunas distribuciones complejas, los momentos pueden no existir o ser difíciles de calcular.
- **Consistencia:** Los estimadores de momentos no siempre son consistentes, especialmente en muestras pequeñas.

- El **método de la máxima verosimilitud** es una técnica estadística ampliamente utilizada para estimar los parámetros desconocidos de una distribución de probabilidad.
- Este método se basa en encontrar los valores de los parámetros que maximicen la función de verosimilitud, la cual mide la probabilidad de observar los datos dados los parámetros. Es decir, la idea básica es seleccionar el valor del parámetro que hace que los datos sean más probables.
- El método de máxima verosimilitud es el método más popular para obtener un estimador.

Dado un modelo estadístico (es decir, una familia de distribuciones $f(\cdot|\theta)$, $\theta \in \Theta$ donde θ es el parámetro del modelo), el método de máxima verosimilitud encuentra el valor del parámetro del modelo θ que maximiza la función de verosimilitud:

$$\hat{\theta}(x) = \max_{\theta \in \Theta} L(\theta|\mathbf{x})$$

Para una muestra aleatoria $\mathbf{x} = (x_1, \dots, x_n)$ de una variable aleatoria X , la **verosimilitud** es proporcional al producto de las probabilidades asociadas a los valores individuales:

$$\prod_j P(X = x_j)$$

Cuando X es una variable aleatoria continua, la verosimilitud es aproximadamente proporcional a

$$\prod_j f(x_j)$$

donde f es la función de densidad de X . Por lo tanto, la **verosimilitud** describe lo plausible que es un valor del parámetro poblacional, dadas unas observaciones concretas de la muestra

Función de Verosimilitud: La función de verosimilitud, $L(\theta|\mathbf{x})$, para un conjunto de datos $\mathbf{x} = (x_1, x_2, \dots, x_n)$ y un vector de parámetros θ , es el producto de las funciones de densidad (o de probabilidad) de los datos observados, dadas las posibles realizaciones de θ :

$$\begin{aligned} L(\theta|\mathbf{x}) &= f(\mathbf{x}|\theta) = f(x_1, \dots, x_n|\theta) = \\ &= f(x_1|\theta)f(x_2|\theta) \dots f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) \end{aligned}$$

donde $f(x_i|\theta)$ es la función de densidad (o de probabilidad) de x_i dado θ .

Log-Verosimilitud: Debido a que la función de verosimilitud puede implicar productos de muchos términos, es más práctico trabajar con su logaritmo natural, conocido como la log-verosimilitud:

$$\ell(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x}) = \sum_{i=1}^n \log f(x_i|\theta)$$

- 1 **Especificar la Función de Verosimilitud:** Identificar la función de verosimilitud correspondiente a los datos observados y a la distribución supuesta.
- 2 **Calcular la Log-Verosimilitud:** Tomar el logaritmo natural de la función de verosimilitud para obtener la función de log-verosimilitud.
- 3 **Derivar y Resolver:** Derivar la función de log-verosimilitud con respecto a cada parámetro y resolver las ecuaciones obtenidas igualando a cero (puntos críticos) para encontrar los estimadores de máxima verosimilitud (EMV).
- 4 **Verificar Máximos:** Asegurarse de que las soluciones encontradas corresponden a máximos y no a mínimos o puntos de inflexión, típicamente verificando la segunda derivada.

¡Veámos cómo obtener estimadores para los parámetros μ y σ^2 de la distribución Normal con el método de la máxima verosimilitud!

Ventajas

- **Consistencia:** Los estimadores de máxima verosimilitud son consistentes
- **Eficiencia:** En muchos casos, los estimadores de máxima verosimilitud son eficientes, alcanzando la varianza mínima entre los estimadores insesgados.
- **Flexibilidad:** Se puede aplicar a una amplia gama de distribuciones y modelos complejos.
- **Invariantes:** Si T es el estimador de máxima verosimilitud para θ , entonces $\tau(T)$ es el estimador de máxima verosimilitud para $\tau(\theta)$ para cualquier función τ .

Limitaciones

- **Complejidad Computacional:** Encontrar los estimadores de máxima verosimilitud puede implicar resolver ecuaciones no lineales, lo cual puede ser complejo y requerir técnicas numéricas.
- **Existencia y Unicidad:** Los estimadores de máxima verosimilitud no siempre existen y, si existen, no siempre son únicos. En problemas reales, la derivada de la función de verosimilitud es, a veces, analíticamente intratable. En esos casos, se utilizan métodos iterativos para encontrar soluciones numéricas para las estimaciones de los parámetros.
- **Sesgo en Muestras Pequeñas:** Los estimadores pueden ser sesgados en muestras pequeñas, aunque el sesgo disminuye a medida que el tamaño de la muestra aumenta.

Gomez Villegas, M. A. (2005). *Inferencia estadística*. Ediciones Díaz de Santos.

Spiegel, M., & Stephens, L. (2009). Estadística–Serie Schaum. *Mc Graw-Hill*.

Wasserman, L. (2013). All of statistics: A concise course in statistical inference.

Canavos, G. C., & Medal, E. G. U. (1987). *Probabilidad y estadística* (p. 651). México: McGraw Hill.