

TEMA 2: INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA.

2.1.- CONCEPTOS FUNDAMENTALES.

Inferencia estadística.

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.

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 materia prima la información muestral. Una clara aplicación se tiene sobre la realidad económica de la que se estudian problemas concretos para mejorar el conocimiento de ciertas poblaciones.

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). En esta documentación se presentan sólo los métodos clásicos (más adecuados para la crítica de modelos) tanto paramétricos como no paramétricos.

Población y muestra.

La **población** se define como conjunto de elementos objeto de estudio (característica de los elementos estudiados). De forma más concreta, si nos referimos a los elementos o unidades elementales o primarias se llega al "Universo" o "Colectivo" (se dice "dado un colectivo de individuos"), si se refiere a la característica que representa la variable aleatoria se tiene la "Población" que se identifica con la variante (se dice "sea una población con comportamiento normal"). Alternativamente, la población se define como conjunto de posibles resultados de un experimento aleatorio. En adelante, se hará un uso "general" de la definición de "población".

El número de elementos u observaciones se denomina tamaño poblacional. Puede ser finito o infinito. Muchas de las aplicaciones o casos reales estudiados son finitos, aunque cuando la población es muy amplia se tratan o asumen como poblaciones infinitas. Como los modelos son idealizaciones matemáticas lo infinito y lo continuo implica simplificación.

La información sobre la población se encuentra en los elementos o unidades elementales o primarias. Se recoge habitualmente en el ámbito económico-social mediante encuestas. Son encuestas censales si se llega a todos los elementos (recuento y observaciones exhaustivas o de todos los individuos lo que conlleva costes muy elevados), sería lo ideal para obtener toda la información. Como no siempre es viable, se necesitan las encuestas muestrales. Son encuestas muestrales si se recoge información sólo de una parte de la población, que representa a toda ella.

La **muestra** se define como un subconjunto representativo de la población. No cualquier parte de una población sirve como muestra, debe ser una parte "representativa" de la población, debe mantener la estructura de la población (en cuanto a la característica estudiada), debe ser una "imagen en pequeño" de la población en la que se mantenga la heterogeneidad o variabilidad existente en la población ("micropoblación").

Si una urna tiene 10.000 bolas de dos colores, la mitad son rojas y la otra mitad azules, una muestra de tamaño 20 debería tener de forma "ideal" 10 rojas y 10 azules. En realidad no son las muestras "ideales" o de "perfecta representatividad" y podría darse una muestra diferente (por ejemplo 9 rojas y 11 azules). De hecho, para una muestra de tamaño "n" se define el espacio de las muestras o "Espacio Muestral" como

el conjunto de todas las muestras (diferentes) de tamaño n que se pueden obtener a partir de una población dada. Las desviaciones respecto de la perfecta representatividad, que se atribuyen al proceso de selección de la muestra, no invalidan el proceso inferencial siempre y cuando tengan origen aleatorio, es decir, sean debidas al azar. El azar se encargará de proporcionar una muestra con la representatividad deseada.

Por limitaciones en los recursos (tiempo acotado, presupuesto o coste limitado, observación destructiva, bienes perecederos) es habitual el uso de muestras en vez de censos. Además, existen ciertas ventajas a tener en cuenta: mayor detalle y calidad en la observación y recogida de datos al ser menor el número de individuos a estudiar, ahorro de tiempo y costes. Los inconvenientes surgen por la incertidumbre inevitable debida al proceso inferencial, por los errores en la forma de seleccionar la muestra, y por la mayor cualificación que se debe tener para efectuar procesos inferenciales.

Según la forma en que se seleccionan las unidades muestrales se distinguen dos grandes **tipos de muestreo**: muestreos no probabilísticos (no estocásticos) y muestreos probabilísticos (estocásticos). A veces se nombran como muestreos no aleatorios y muestreos aleatorios. Ésta última denominación, aunque muy extendida, es menos apropiada ya que aleatorio se refiere propiamente a azar o ambiente de riesgo (no necesariamente evaluable o medible) y probabilístico o estocástico indica posibilidad de asignar probabilidades en un ambiente de incertidumbre.

En los muestreos no probabilísticos el investigador selecciona determinísticamente las unidades muestrales (o al menos no es posible asignar probabilidades). ¿Podrían conseguirse muestras o subconjuntos representativos de esta forma? En principio, sí. ¿Podrían conseguirse buenas estimaciones o estimaciones cercanas a los valores poblacionales? En principio, sí. El problema se tiene al no poder calcular ninguna medida de bondad o calidad para la estimación. Al no inscribirse en un entorno matemático este tipo de muestreo presenta el grave inconveniente de que no permite el análisis estadístico (valoración de seguridad y precisión de conclusiones, estudio de errores cometidos en la estimación, fijación de tamaños muestrales para obtener objetivos, etc.). Esto es, no se tiene una objetivación del procedimiento. Ejemplos de muestreos no probabilísticos son el muestreo opinático (según criterio del investigador) o el muestreo errático (sin criterio).

Los muestreos probabilísticos se caracterizan porque cada elemento de la población tiene una probabilidad conocida y no nula de ser seleccionado. Por ello, mediante técnicas estadísticas se tienen los resultados que se valoran, pudiéndose medir y controlar los errores.

En los muestreos probabilísticos cada elemento muestral (a priori o al plantear el mecanismo de selección) es una variante (con posibles valores y probabilidades asociadas). Como cada elemento es una variante, la muestra en su conjunto es una variable aleatoria (con espacio muestral y probabilidades asociadas).

Existen diferentes tipos de muestreos probabilísticos. Se comentan el muestreo sin reemplazamiento (también denominado muestreo aleatorio sin reemplazamiento o muestreo aleatorio sin reemplazamiento) y el muestreo aleatorio simple o muestreo irrestrictamente aleatorio. La diferencia entre ambos viene del no reemplazamiento o reemplazamiento (equivalente a población infinita) tras la extracción de los elementos muestrales. Esto implica dependencia o independencia entre las variables aleatorias representativas de los elementos muestrales a extraer.

El muestreo sin reemplazamiento (también denominado muestreo aleatorio sin reemplazamiento o muestreo aleatorio sin reemplazamiento) sólo exige que cada elemento muestral tenga igual comportamiento que el poblacional y que todas las muestras de tamaño n tengan la misma probabilidad de ser escogidas. En general, para que la muestra sea representativa, la elección de los elementos de la población de los que se tomará información sobre la característica de interés debe de hacerse en condiciones de azar.

Un **ejemplo** de muestreo sin reemplazamiento es extraer de una urna con 100 bolas (20 marcadas con el número uno, 30 marcadas con el número dos y 50 marcadas con el número tres) dos bolas al azar, sabiendo que no se devuelve la primera bola extraída (la primera se elige entre una urna con 100 bolas y la segunda se elige entre una urna con 99 bolas). La primera extracción influye o provoca dependencia en la segunda ya que las probabilidades cambian (la población en la primera extracción es de 100 bolas y luego, en la segunda extracción, la población cambia ya que es de 99 bolas).

La población inicial es:

$\xi=x_i$	$P(\xi=x_i)$
1	0,2
2	0,3
3	0,5
Total	1

A priori, la probabilidad de que la primera extracción obtenga 1, 2 o 3 es:

$$P[(X_1=1 \cap X_2=1) \cup (X_1=1 \cap X_2=2) \cup (X_1=1 \cap X_2=3)] = \frac{20}{100} \cdot \frac{19}{99} + \frac{20}{100} \cdot \frac{30}{99} + \frac{20}{100} \cdot \frac{50}{99} = 0,2$$

$$P[(X_1=2 \cap X_2=1) \cup (X_1=2 \cap X_2=2) \cup (X_1=2 \cap X_2=3)] = \frac{30}{100} \cdot \frac{20}{99} + \frac{30}{100} \cdot \frac{29}{99} + \frac{30}{100} \cdot \frac{50}{99} = 0,3$$

$$P[(X_1=3 \cap X_2=1) \cup (X_1=3 \cap X_2=2) \cup (X_1=3 \cap X_2=3)] = \frac{50}{100} \cdot \frac{20}{99} + \frac{50}{100} \cdot \frac{30}{99} + \frac{50}{100} \cdot \frac{49}{99} = 0,5$$

A priori, la probabilidad de que la segunda extracción obtenga 1, 2 o 3 es:

$$P[(X_1=1 \cap X_2=1) \cup (X_1=2 \cap X_2=1) \cup (X_1=3 \cap X_2=1)] = \frac{20}{100} \cdot \frac{19}{99} + \frac{30}{100} \cdot \frac{20}{99} + \frac{50}{100} \cdot \frac{20}{99} = 0,2$$

$$P[(X_1=1 \cap X_2=2) \cup (X_1=2 \cap X_2=2) \cup (X_1=3 \cap X_2=2)] = \frac{20}{100} \cdot \frac{30}{99} + \frac{30}{100} \cdot \frac{29}{99} + \frac{50}{100} \cdot \frac{30}{99} = 0,3$$

$$P[(X_1=1 \cap X_2=3) \cup (X_1=2 \cap X_2=3) \cup (X_1=3 \cap X_2=3)] = \frac{20}{100} \cdot \frac{50}{99} + \frac{30}{100} \cdot \frac{50}{99} + \frac{50}{100} \cdot \frac{49}{99} = 0,5$$

Ambos elementos muestrales se distribuyen como la población pero la probabilidad de cada muestra concreta (probabilidades conjuntas) no son el producto de las marginales dado que no hay independencia, el caso de la muestra (1,1) sería:

$$P(X_1=1 \cap X_2=1) \neq P(X_1=1) \cdot P(X_2=1)$$

$$\frac{20}{100} \cdot \frac{19}{99} \neq \frac{20}{100} \cdot \frac{20}{100}$$

Entre los muestreos probabilísticos se estudiará y se utilizará en adelante el m.a.s. o muestreo aleatorio simple o muestreo irrestrictamente aleatorio.

Muestreo aleatorio simple o muestreo irrestrictamente aleatorio. Es un método de selección de un número fijo de unidades muestrales n , que garantiza que todas las muestras de tamaño n tienen la misma probabilidad de ser escogidas. Además,

los elementos muestrales serán variantes independientes entre sí y con igual distribución de probabilidad que la población (extracciones con repetición o en población infinita), esa independencia e igual distribución se puede indicar como ĩ.i.d.ö. Es importante observar que, en todo caso, la muestra será una variable aleatoria antes de su concreción (a priori) y será un valor o concreción de la variable aleatoria tras su obtención (a posteriori). Como variable aleatoria tendrá su distribución de probabilidad determinada por la población de origen y por el tipo de muestreo.

Un **ejemplo** de muestreo con reemplazamiento es extraer de una urna con 100 bolas (20 marcadas con el número ũunoö, 30 marcadas con el número ödosö y 50 marcadas con el número ötresö) dos bolas al azar, sabiendo que se ödevuelveö la primera bola extraída (la primera se elige entre una urna con 100 bolas y la segunda se elige también entre una urna con 100 bolas). La primera extracción no influye o no provoca dependencia en la segunda ya que las probabilidades no cambian (la población en la primera extracción es de 100 bolas y luego, en la segunda extracción, la población sigue igual).

La población inicial es:

$\xi=x_i$	$P(\xi=x_i)$
1	0,2
2	0,3
3	0,5
Total	1

A priori, la probabilidad de que la primera extracción obtenga 1, 2 o 3 es:

$$P[(X_1=1 \cap X_2=1) \cup (X_1=1 \cap X_2=2) \cup (X_1=1 \cap X_2=3)] = 0,2 \cdot 0,2 + 0,2 \cdot 0,3 + 0,2 \cdot 0,5 = 0,2$$

$$P[(X_1=2 \cap X_2=1) \cup (X_1=2 \cap X_2=2) \cup (X_1=2 \cap X_2=3)] = 0,3 \cdot 0,2 + 0,3 \cdot 0,3 + 0,3 \cdot 0,5 = 0,3$$

$$P[(X_1=3 \cap X_2=1) \cup (X_1=3 \cap X_2=2) \cup (X_1=3 \cap X_2=3)] = 0,5 \cdot 0,2 + 0,5 \cdot 0,3 + 0,5 \cdot 0,5 = 0,5$$

A priori, la probabilidad de que la segunda extracción obtenga 1, 2 o 3 es:

$$P[(X_1=1 \cap X_2=1) \cup (X_1=2 \cap X_2=1) \cup (X_1=3 \cap X_2=1)] = 0,2 \cdot 0,2 + 0,3 \cdot 0,2 + 0,5 \cdot 0,2 = 0,2$$

$$P[(X_1=1 \cap X_2=2) \cup (X_1=2 \cap X_2=2) \cup (X_1=3 \cap X_2=2)] = 0,2 \cdot 0,3 + 0,3 \cdot 0,3 + 0,5 \cdot 0,3 = 0,3$$

$$P[(X_1=1 \cap X_2=3) \cup (X_1=2 \cap X_2=3) \cup (X_1=3 \cap X_2=3)] = 0,2 \cdot 0,5 + 0,3 \cdot 0,5 + 0,5 \cdot 0,5 = 0,5$$

Ambos elementos muestrales se distribuyen como la población y la probabilidad de cada muestra concreta (probabilidades conjuntas) son el producto de las marginales dado que hay independencia:

$$P(X_1 = x_1 \cap X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$$

Parámetro y estimador.

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*; 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.

Un Estadístico es cualquier función muestral (no incluye parámetros desconocidos). Ejemplos pueden ser la media aritmética, la varianza o la proporción muestral. Como cada elemento muestral (a priori o al plantear el mecanismo de selección) es una variante (con posibles valores y probabilidades asociadas) y como la muestra en su conjunto es una variable aleatoria (con espacio muestral y probabilidades asociadas), se llega a que cualquier estadístico es, también, una variante.

El Estimador será un Estadístico dedicado al conocimiento de un parámetro poblacional desconocido. Por tanto, 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 y la varianza muestral sean estadísticos relevantes cuya distribución de probabilidad en el muestreo necesita ser conocida por su amplia utilización. Estos estadísticos a los que se les confiere la característica de ser útiles para estimar se denominan estimadores.

Resumiendo, dado un tipo de muestreo probabilístico a partir del cual se obtiene una muestra, se denomina estimador a una función de la muestra $\hat{\theta}(x_1, \dots, x_n) = \hat{\theta}$ que se usa para inferir el valor de la característica poblacional que queremos estimar θ . El conjunto de valores que puede tomar el parámetro se denomina espacio paramétrico Θ .

2.2.- DISTRIBUCIÓN DE ESTADÍSTICOS EN EL MUESTREO.

El **estadístico** es una función de la muestra que no presenta parámetros poblacionales desconocidos: $T(X) = f(X_1, X_2, \dots, X_n)$.

Algunos ejemplos de estadísticos son el total muestral, la media muestral, el momento de orden tres respecto al origen o el menor valor muestral:

$$f(X_1, X_2, \dots, X_n) = X_1 + X_2 + \dots + X_n$$

$$f(X_1, X_2, \dots, X_n) = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$f(X_1, X_2, \dots, X_n) = \frac{X_1^3 + X_2^3 + \dots + X_n^3}{n}$$

$$f(X_1, X_2, \dots, X_n) = \min(X_1; X_2; \dots; X_n)$$

Es importante destacar que el estadístico presenta una **distribución de probabilidad derivada de la muestral** (definida según población de origen y tipo de muestreo) y de la forma de la función que sigue el propio estimador. Además, consecuencia de lo anterior, supone o suministra un resumen de la información contenida en la muestra. La distribución de la muestra suele identificarse con su función de cuantía o de densidad conjunta (según que la distribución poblacional sea discreta o continua), que proporcionan la probabilidad o densidad de probabilidad con la cual puede presentarse cada muestra concreta en el proceso de muestreo.

En este curso siempre se tiene que el tipo de muestreo es **m.a.s.** Por ello se recuerdan y analizan sus características. Es un método de selección de un número fijo de unidades muestrales n , que garantiza que todas las muestras de tamaño n tienen la misma probabilidad de ser escogidas. Además, los elementos muestrales serán variantes independientes entre sí y con igual distribución de probabilidad que la población (extracciones con repetición o en población infinita), esa independencia e igual distribución se puede indicar como *ö.i.d.ö*. Es importante recordar que la muestra será una variable aleatoria antes de su concreción (a priori), se suele representar con letras mayúsculas que indican las variantes elementos muestrales $X(X_1, X_2, \dots, X_n)$. Es importante recordar que la muestra será un valor o concreción de la variable aleatoria tras su obtención (a posteriori), se suele representar con letras minúsculas que indican las concreciones o sucesos muestrales $x(x_1, x_2, \dots, x_n)$. Como variable aleatoria tendrá su distribución de probabilidad determinada por la población de origen y por el tipo de muestreo que es m.a.s.

Para una muestra, m.a.s. $x(x_1, x_2, \dots, x_n)$, su probabilidad es si caso discreto:

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

.

Para una muestra, m.a.s. $x(x_1, x_2, \dots, x_n)$, su densidad de probabilidad es si caso continuo:

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

Hay que recordar la independencia (se pasa de intersección de sucesos a producto de probabilidades) y la igual distribución (se pasa de variante elemento muestral a variante poblacional X_j).

Un **ejemplo** es extraer de una población $B(1, p)$ una m.a.s. de tamaño tres ($n=3$) y obtener la distribución de probabilidad de la muestra:

$X_1=x_1; X_2=x_2; X_3=x_3$	$P(X_1=x_1; X_2=x_2; X_3=x_3)$
1; 1; 1	p^3
1; 1; 0	p^2q
1; 0; 1	p^2q
0; 1; 1	p^2q
1; 0; 0	pq^2
0; 1; 0	pq^2
0; 0; 1	pq^2
0; 0; 0	q^3

Se observa en la columna de la izquierda el espacio muestral con ocho diferentes valores de dimensión tres. Se observa en la columna de la derecha las probabilidades asociadas a los diferentes valores posibles de la muestra, se comprueba que su suma es uno (toda la probabilidad del fenómeno o suceso seguro).

Otro **ejemplo** es extraer de una población normal una m.a.s. de tamaño n y buscar la función de densidad conjunta de la muestra:

$$\begin{aligned}
 f(x_1, x_2, \dots, x_n) &= f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) \\
 &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x_2-\mu)^2}{\sigma^2}} \cdot \dots \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x_n-\mu)^2}{\sigma^2}} \\
 &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2}} \quad -\infty < x_i < \infty
 \end{aligned}$$

Otro **ejemplo** es extraer de una urna con 100 bolas (20 marcadas con el número uno, 30 marcadas con el número dos y 50 marcadas con el número tres) una m.a.s. de tamaño dos y hallar distribución poblacional, distribución de los elementos muestrales, distribución conjunta de la muestra y distribución de la media muestral.

$\xi=x_i$	$P(\xi=x_i)$
1	0,2
2	0,3
3	0,5
Total	1

$X_1=x_1$	$P(X_1=x_1)$
1	0,2
2	0,3
3	0,5
Total	1

$X_2=x_2$	$P(X_2=x_2)$
1	0,2
2	0,3
3	0,5
Total	1

$X=x$	$P(X=x)=P(X_1=x_1) \cdot P(X_2=x_2)$	Media
(1,1)	$0,2 \cdot 0,2 = 0,04$	1
(1,2)	$0,2 \cdot 0,3 = 0,06$	1,5

$(X_1+X_2)/2$	$P[\text{Media}]$
1	0,04
1,5	0,12

(2,1)	$0,3 \cdot 0,2 = 0,06$	1,5
(2,2)	$0,3 \cdot 0,3 = 0,09$	2
(1,3)	$0,2 \cdot 0,5 = 0,1$	2
(3,1)	$0,5 \cdot 0,2 = 0,1$	2
(2,3)	$0,3 \cdot 0,5 = 0,15$	2,5
(3,2)	$0,5 \cdot 0,3 = 0,15$	2,5
(3,3)	$0,5 \cdot 0,5 = 0,25$	3
Total	1	

2	0,29
2,5	0,3
3	0,25
Total	1

Con este ejemplo se ðcomprendeð que cualquier estadístico (como la media) es una variante con posibles valores y probabilidades asociadas (con distribución de probabilidad). También se comprende que supone un ðresumenð del espacio muestral: en el ejemplo este espacio muestral es de dimensión dos con nueve valores diferentes mientras que la media es de dimensión uno con cinco valores diferentes.

DISTRIBUCIONES DE ESTADÍSTICOS (SUPUESTO M.A.S.).

Población desconocida pero con media y varianzas conocidas $E[\xi] = \mu$ y $Var[\xi] = \sigma^2$.

Sea $X(X_1, X_2, \dots, X_n)$ una muestra aleatoria (m.a.s.). Entonces las variables aleatorias que representan cada elemento muestral están idénticamente distribuidas que la población con $E[X_i] = \mu$ y $Var[X_i] = \sigma^2$. Además son independientes dado el tipo de muestreo. Si no se conoce la distribución de la población, no se puede, en general, calcular la distribución de los estadísticos, pero sí se podrá, en cualquier caso, determinar la esperanza y varianzas de la media muestral y de la varianzas muestral.

Variable aleatoria ðMedia Muestralð $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ con

esperanza de la media muestral $E[\bar{X}] = E[\xi] = \mu$ ya que

$$E(\bar{X}) = \frac{1}{n} E \left[\sum_{i=1}^n X_i \right] = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n E(\xi) = E(\xi)$$

Varianza de la media muestral $Var[\bar{X}] = \frac{\sigma^2}{n}$ ya que

$$V(\bar{X}) = \frac{1}{n^2} V\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} n V(\xi) = \frac{\sigma^2}{n}$$

Además, siempre que n sea muy grande ($n > 30$) podemos aplicar el TCL:

$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Variable aleatoria òVarianza muestralö $s^2 = \frac{\sum (X_i - \bar{X}_n)^2}{n} = \frac{\sum X_i^2}{n} - \bar{X}_n^2$

con media de la Varianza muestral $E[s^2] = \frac{n-1}{n} \sigma^2$ y

Con varianza de la Varianza muestral $Var[s^2] = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} - \frac{\mu_4 - 3\mu_2^2}{n^3}$

(llamando $\mu_n = m_n = E[\xi - E(\xi)]^n$, entonces $\mu_2 = m_2 = V(\xi) = \sigma^2$).

Si población normal $\mu_4 = 3\mu_2^2 = 3\sigma^4$ entonces $Var[s^2] = \frac{2(n-1)\sigma^4}{n^2}$

Población Normal

Las distribuciones de mayor interés son las referidas a m.a.s. sobre distribuciones normales. En ciertos casos, por el T.C.L., serían aproximaciones para el caso de grandes muestras.

Para **una muestra**. Sean $X(X_1, X_2, \dots, X_n)$ una muestra aleatoria simple de variables aleatorias tales que $\xi \equiv X_i \rightarrow N(\mu, \sigma) \forall i$

Variable aleatoria òMedia Muestralö.

$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \text{ o } \bar{X} \rightarrow \frac{\sigma}{\sqrt{n}} N(0,1) + \mu$$

Se justifica con la propiedad aditiva de la normal y con los resultados anteriores

$$E[\bar{X}] = \mu \quad \text{y} \quad \text{Var}[\bar{X}] = \frac{\sigma^2}{n} .$$

Otra expresión para la misma idea puede ser que por m.a.s. $X_i \rightarrow N(\mu, \sigma^2)$, por el teorema de la adición se cumple que $X_1 + X_2 + \dots + X_n \rightarrow N(n\mu, n\sigma^2)$, y por las transformaciones lineales $\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Si no se conoce la varianza poblacional no es operativa la expresión

$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) , \text{ o equivalente } \bar{X} \rightarrow \frac{\sigma}{\sqrt{n}} N(0,1) + \mu .$$

Si no se conoce el valor de σ^2 , se puede utilizar otra expresión que introduce la varianza muestral en vez de la poblacional, entonces:

$$\frac{\bar{X} - \mu}{s} \sqrt{n-1} \rightarrow t_{n-1} , \text{ o equivalente } \bar{X} \rightarrow \frac{s}{\sqrt{n-1}} t_{n-1} + \mu , \text{ o}$$

$$\bar{X} \rightarrow \frac{s_{n-1}}{\sqrt{n}} t_{n-1} + \mu \quad (\text{utilizando la cuasivarianza muestral}).$$

Variable aleatoria ðVarianza muestralö.

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{\sum X_i^2}{n} - \bar{X}^2$$

$$\frac{ns^2}{\sigma^2} \rightarrow \chi_{n-1}^2 \quad ; \quad E[s^2] = \frac{n-1}{n} \sigma^2 \quad ; \quad \text{Var}[s^2] = \frac{2(n-1)\sigma^4}{n^2} .$$

Se prescinde de la información sobre la media poblacional.

Población Binomial.

Variable aleatoria ðProporción Muestralö.

Se tiene población Bernoulli y, entonces, $\xi \equiv X_i \rightarrow B(1; p)$.

Se considera la variable aleatoria $X =$ "nº de éxitos en la muestra", que sigue una distribución $B(n, p)$. Se define $P =$ "proporción de éxitos en n extracciones", entonces,

$$P = X/n = B(n, p)/n \quad ; \quad E[P] = p \quad ; \quad \text{Var}[P] = \frac{pq}{n} .$$

La proporción muestral es una media en la que los valores son o cero o uno. Aunque la distribución más correcta es la proporción binomial, por el TCL se puede aproximar si la muestra es grande a

$$\bar{X} \rightarrow P \rightarrow N\left(p, \sqrt{\frac{pq}{n}}\right).$$

Dos poblaciones normales (muestras independientes).

Para **dos muestras independientes** se supone que se quiere comparar dos variables o dos poblaciones. Se toman dos muestras y se suponen independientes. Se sigue suponiendo que la distribución de las dos poblaciones es Normal.

Sean X_1, X_2, \dots, X_n una muestra aleatoria de variables aleatorias independientes tales que $X_i \rightarrow N(\mu_x, \sigma_x) \forall i = 1, 2, \dots, n$ e Y_1, Y_2, \dots, Y_m una muestra aleatoria de variables aleatorias tales que $Y_j \rightarrow N(\mu_y, \sigma_y) \forall j = 1, 2, \dots, m$.

Variante Diferencia de medias muestrales (varianzas poblacionales conocidas).

$$\bar{X}_n - \bar{Y}_m \rightarrow N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right)$$

Un caso particular sería la diferencia de proporciones.

$$P_1 - P_2 \rightarrow N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}\right)$$

Variante Diferencia de medias muestrales (varianzas poblacionales desconocidas pero iguales).

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_x - \mu_y)}{\sqrt{\frac{ns_x^2 + ms_y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m}\right)}} \rightarrow t_{n+m-2}$$

Variante Cociente de varianzas muestrales.

$$\frac{\frac{n \cdot s_x^2}{\sigma_x^2 (n-1)}}{\frac{m \cdot s_y^2}{\sigma_y^2 (m-1)}} \rightarrow F_{(n-1), (m-1)}$$

Dos poblaciones normales (muestras relacionadas, datos apareados).

Se tienen **dos muestras relacionadas** para comparar dos variables, esto es, las dos muestras que se toman no son independientes (los datos están apareados).

Sean X_1, X_2, \dots, X_n una muestra aleatoria de variables aleatorias tales que $X_i \rightarrow N(\mu_x, \sigma_x) \forall i = 1, 2, \dots, n$ e Y_1, Y_2, \dots, Y_n una muestra aleatoria de variables aleatorias tales que $Y_j \rightarrow N(\mu_y, \sigma_y) \forall j = 1, 2, \dots, n$. Las muestras están relacionadas. Notar que en este caso $n=m$.

Variante Diferencia de medias muestrales.

Sea $D_i = X_i - Y_i$ y sea s_d su desviación típica muestral, se puede usar el siguiente resultado:

$$\sqrt{n-1} \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{s_d} \rightarrow t_{n-1}$$

Tras observar estas distribuciones, el siguiente paso está en el análisis de los posibles estadísticos estimadores. Será muy importante establecer las propiedades deseables en el estimador para poder juzgar, inicialmente, su bondad, ya que siempre permanecerá desconocido el verdadero valor del parámetro poblacional. Además, posteriormente (tema 3) conviene establecer métodos de obtención de estimadores ya que se podrían plantear, en principio, una gran cantidad de estadísticos estimadores sobre los que habría que comprobar sus propiedades.

2.3.- PROPIEDADES DE LOS ESTIMADORES.

Para estudiar la bondad de un estimador se puede empezar analizando el error cometido en la estimación. Éste se define como la distancia entre la estimación realizada y el verdadero valor del parámetro $\hat{\theta} - \theta$. Este valor siempre será desconocido dado que

el valor del parámetro no se conocerá y que el estimador es una variable aleatoria. Para obtener una medida más operativa se procede como sigue: se elimina la influencia del signo en la diferencia con el cuadrado $(\hat{\theta} - \theta)^2$ y se toma la esperanza. Con ello, se llega a una medida denominada error cuadrático medio (representa un criterio intuitivo y operativo). Se define el **error cuadrático medio (ECM)** como $ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2$.

Un valor pequeño del error cuadrático medio significa que el estimador elegido, en media, no se encuentra lejos del parámetro desconocido (este podría ser un buen criterio para comparar distintos estimadores). Se puede plantear elegir el estimador que minimice el error cuadrático medio.

Estudiando la expresión anterior se puede llegar a $ECM(\hat{\theta}) = V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$.

EJEMPLO.

Descomponer el error cuadrático medio.

$$\begin{aligned} ECM(\hat{\theta}) &= E((\hat{\theta} - \theta)^2) = E((\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta))^2 = \\ &= E\left((\hat{\theta} - E(\hat{\theta}))^2 + 2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2\right) = \\ &= E\left((\hat{\theta} - E(\hat{\theta}))^2\right) + 2E(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2 = \\ &= V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 \end{aligned}$$

puesto que $E(\hat{\theta} - E(\hat{\theta})) = E(\hat{\theta}) - E(\hat{\theta}) = 0$ y $E((E(\hat{\theta}) - \theta)^2) = (E(\hat{\theta}) - \theta)^2$ por ser la esperanza de una constante.

La descomposición del E.C.M. en dos sumandos positivos indica que para encontrar un buen estimador es necesario minimizar ambos simultáneamente. En general, esto no es posible (basta observar que la varianza mínima se obtendría a través de estimadores constantes para los que no tenemos ningún control sobre el otro sumando). Una solución es restringir la búsqueda de estimadores a una clase especial de los mismos, los estimadores insesgados.

Estimador insesgado.

Un estimador es insesgado cuando su media o esperanza matemática coincide con el valor verdadero del parámetro desconocido, esto es, $E(\hat{\theta}) = \theta$

Si el estimador es insesgado, minimiza el segundo sumando en que se descompone el error cuadrático medio, $(E(\hat{\theta}) - \theta)^2 = 0$ y, por tanto, el error cuadrático medio coincide con la varianza del estimador.

Al ser insesgado, el valor del parámetro desconocido es la media de los valores de $\hat{\theta}$ para todas las muestras posibles (con esta propiedad se consigue que la estimación sea insesgada en término medio).

Si un estimador no es insesgado se denomina sesgo a la diferencia $b(\theta) = E(\hat{\theta}) - \theta$. Entre dos estimadores sesgados será mejor aquel cuyo sesgo sea menor en valor absoluto.

Un estimador es asintóticamente insesgado si su sesgo tiende a cero cuando crece el tamaño muestral (tamaño tiende a infinito).

Como propiedad se puede destacar que si existen dos estimadores con el mismo sesgo se pueden construir infinitos estimadores de esa clase mediante combinación lineal convexa de los dos estimadores iniciales.

EJERCICIO.

Demostrar la propiedad anterior.

EJERCICIO.

Demostrar que con m.a.s. cualquier combinación lineal convexa de los elementos muestrales es un estimador insesgado de la media poblacional.

Estimador eficiente.

La eficiencia se relaciona con la minimización de la varianza del estimador (se relaciona con el primer sumando del criterio del error cuadrático medio). Se estudia, especialmente, en el caso de estimadores insesgados (en este caso el error cuadrático medio coincide con la varianza del estimador). Entonces, se llega al estimador que tiene mínima varianza entre todos los estimadores insesgados.

En principio, para analizar la eficiencia de los estimadores bastaría con hallar las varianzas y compararlas. El problema es mayor cuando existen muchos estimadores

en estudio o cuando interesa el estimador con menor varianza de todos los posibles. Se recurre a la información que proporciona la función de verosimilitud mediante la Cota de Cramer-Rao (CCR):

$$V(\hat{\theta}) \leq CCR = \frac{[1 + b'(\theta)]^2}{E\left[\frac{\partial \ln L(X, \theta)}{\partial \theta}\right]^2} = \frac{[1 + b'(\theta)]^2}{I[\theta]} \quad \text{con} \quad b'(\theta) = \frac{\partial b(\theta)}{\partial \theta}.$$

Si se toman estimadores insesgados y muestras aleatorias simples se llega a:

$$V(\hat{\theta}) \leq CCR = \frac{1}{nE\left[\frac{\partial \ln f(x, \theta)}{\partial \theta}\right]^2} = \frac{1}{I[\theta]}.$$

La medida de eficiencia queda como $ME = \frac{CCR}{V(\hat{\theta})}$ con $0 \leq ME \leq 1$.

Se tiene un estimador más eficiente cuanto más cerca de uno está la medida. Si $CCR=1$ es eficiente. Para grandes muestras interesa que sean estimadores asintóticamente eficientes.

Estimador consistente.

La muestra proporciona información sobre los parámetros poblacionales en estudio. La cantidad de información disponible aumenta con el tamaño de la muestra. Un estimador razonable debería recoger ese aumento de información de modo que las estimaciones que se realicen con él sean tanto mejores cuanto mayor sea el número de unidades observadas. En el límite, cuando la muestra coincide con la población la estimación debería coincidir con el verdadero valor del parámetro. Por esto, conviene estudiar el comportamiento de los estimadores en función del tamaño muestral.

Interesa, especialmente, la consistencia en probabilidad. Un estimador es consistente si la probabilidad de que la desviación entre el estimador y el valor verdadero del parámetro sea superior a cualquier número ε por pequeño que sea, se acerca a cero cuando el número de elementos de la muestra se acerca al número de elementos en la población (a infinito en poblaciones infinitas).

$$P(|\hat{\theta} - \theta| > \varepsilon) \xrightarrow{n \rightarrow N \rightarrow \infty} 0, \text{ o bien, } P(|\hat{\theta} - \theta| < \varepsilon) \xrightarrow{n \rightarrow N \rightarrow \infty} 1.$$

Un estimador $\hat{\theta}$ será por tanto consistente si cuando se observa toda la población la estimación coincide exactamente con el valor del parámetro a estimar. En esta situación, al incrementar el tamaño muestral hasta N la muestra coincide con la población y el error cuadrático será cero.

Para poblaciones infinitas se aplica la desigualdad de Tchebychev y se llega a que la condición anterior equivale a una situación de estimador asintóticamente insesgado y con varianza asintóticamente nula.

En este curso se admitirá siempre la hipótesis de **identificar consistencia con estimador asintóticamente insesgado y con varianza asintóticamente nula.**

EJERCICIO.

Demostrar que la consistencia en probabilidad equivale a estimador asintóticamente insesgado y con varianza asintóticamente nula.

Estimador suficiente.

Un estadístico es suficiente cuando contiene toda la información relevante contenida en la muestra respecto del parámetro desconocido (ningún otro estadístico puede proporcionar información adicional sobre el parámetro poblacional desconocido).

Si el estimador es suficiente la distribución de la muestra condicionado al estimador es independiente del parámetro. El procedimiento más operativo para demostrar suficiencia se tiene aplicando el criterio de factorización de Fisher-Neyman.

Estimador robusto.

Un estimador es robusto cuando los cambios en las hipótesis de comportamiento poblacional (modelo poblacional) no afectan o afectan débilmente al estimador.