INFERENCIA ESTADÍSTICA

CONCEPTOS BÁSICOS

Población

En el contexto de la estadística, una población es el conjunto de todos los valores que puede tomar una característica medible en particular, de un conjunto correspondiente de entes reales o abstractos.

A manera de ejemplo, si se está interesado en el diámetro de los lápices que se producen en una fábrica, la población es la característica medible "diámetro" y está constituida por todos las diámetros que miden los lápices que se producen, y no por los lápices en sí.

La forma en que se distribuye la característica de interés en la población se denomina ***distribución poblacional***.

Parámetro poblacional

Es una función de los elementos de la población.

Ejemplos de parámetros son la media y la varianza si son calculados con base en toda la información de la población. Los parámetros a menudo caracterizan a la población. Suelen ser representados con letras griegas minúsculas.

$$\mu \qquad \sigma^2 \qquad \lambda \qquad \theta$$

Para una población con distribución normal, sus parámetros son la media y la varianza, que suelen representarse con μ y σ^2 respectivamente.

Para una población con distribución exponencial, el parámetro suele representarse con λ. *La media de una VA exponencial es* $1/\lambda$.

Muestra

Una muestra es cualquier subconjunto de observaciones de la característica de interés tomadas de la población.

$$\underline{X}=\{X_1,\ X_2,\ X_3,\ \ldots X_{i},\ldots,\ X_n\}$$

Al proceso de seleccionar la muestra se le demonina: muestreo.

El muestreo puede ser aleatorio o no aleatorio; con reemplazo o sin reemplazo.

Los métodos de la Estadística Inferencial requieren que el muestreo sea aleatorio.

Muestra aleatoria

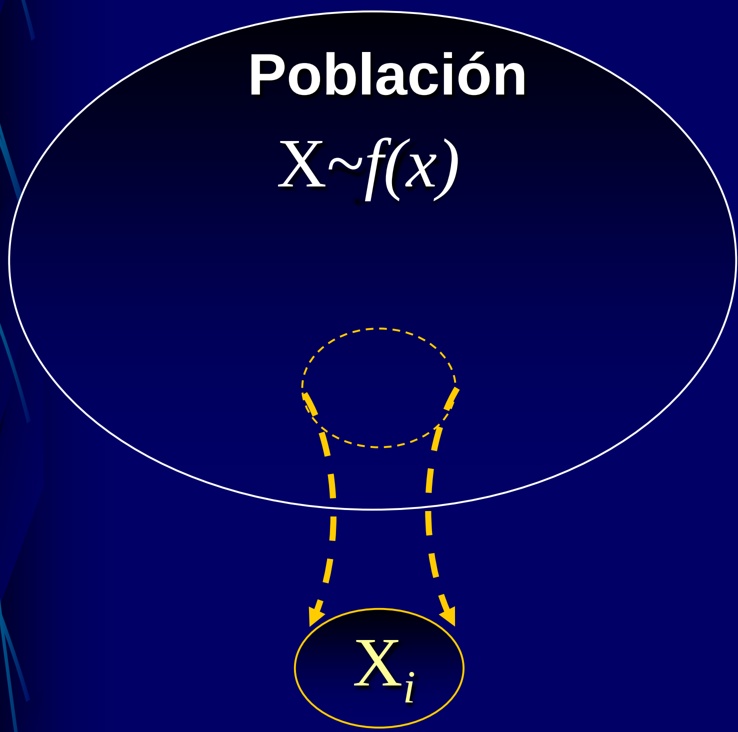
Sea una población con la característica medible X cuya distribución es $f(x)$. Se dice que $\underline{X}=\{x_1, x_2, x_3, \ldots, x_n\}$ es una **muestra aleatoria simple** si:

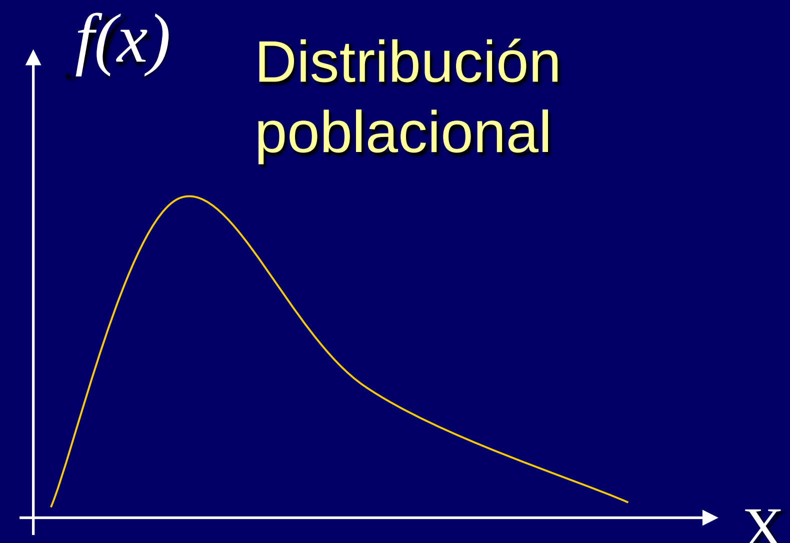
a) Cada valor x_i se obtiene observando X de manera independiente bajo las misma condiciones n veces, es decir, las x_i son variables aleatorias independientes (*lo que significa que se ha efectuado un muestreo aleatorio con reemplazo*).

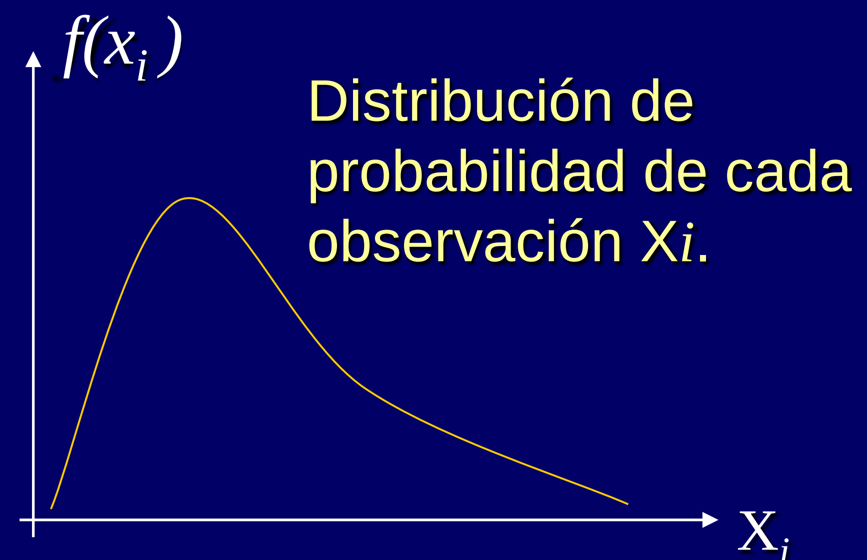
b) Cada observación x_i es una VA cuya distribución de probabilidad es idéntica a la distribución de la población:

$$f(x_i) = f(x) \,, \quad \forall\ i = 1,2,\ldots,n$$

DISTRIBUCIÓN DE CADA OBSERVACIÓN DE LA MUESTRA

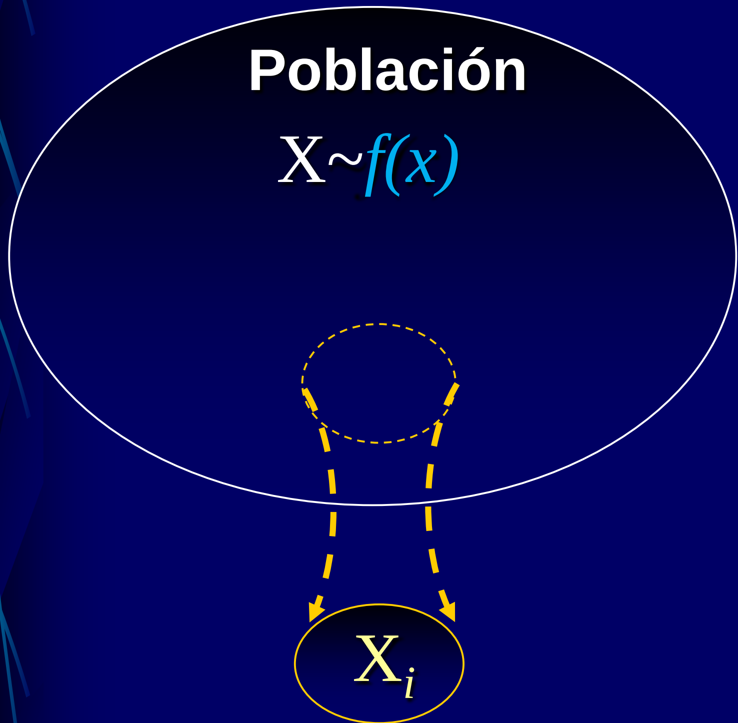






Al extraer una observación de la población, el resultado es una VA cuya distribución de probabilidad es idéntica a la distribución poblacional.

DISTRIBUCIÓN DE PROBABILIDAD PARA CADA OBSERVACIÓN DE LA MUESTRA



Si la media poblacional es

$$\mu$$

Si la varianza poblacional es

$$\sigma^2$$

y se cumplen las condiciones del muestreo aleatorio simple

entonces:

El valor esperado y la varianza de cada VA X_i son iguales a los de la población, ya que $f(x_i)=f(x)$

$$E[X_i]=\mu \qquad V[X_i]=\sigma^2$$

Resultando que la distribución de probabilidad *f(xi)* para todas las Xi de la muestra, es idéntica a la distribución poblacional *f(x)*, con los mismos parámetros.

La muestra está compuesta por n observaciones X_i

La distribución de probabilidad conjunta de las n observaciones X_i que conforman una muestra aleatoria simple es:

$$L(x_1, x_2, x_3, \ldots, x_n) = f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot \ldots \cdot f(x_n)$$

y se conoce como

función de verosimilitud de la muestra.

$$L(\underline{X}) = \prod_{i=1}^{n} f(x_i)$$

Todas las muestras de tamaño n tienen la misma probabilidad de ser seleccionadas

Estadístico

Es una función de las observaciones de la muestra y que no contiene cantidades desconocidas.

Un ejemplo de un *estadístico* es el total muestral:

$$T = \sum_{i=1}^{n} x_i$$

Otro ejemplo de un *estadístico* es la media muestral:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Generalmente los estadísticos se representan con letras latinas.

EJEMPLO

Sea una población compuesta por 5 calificaciones independientes

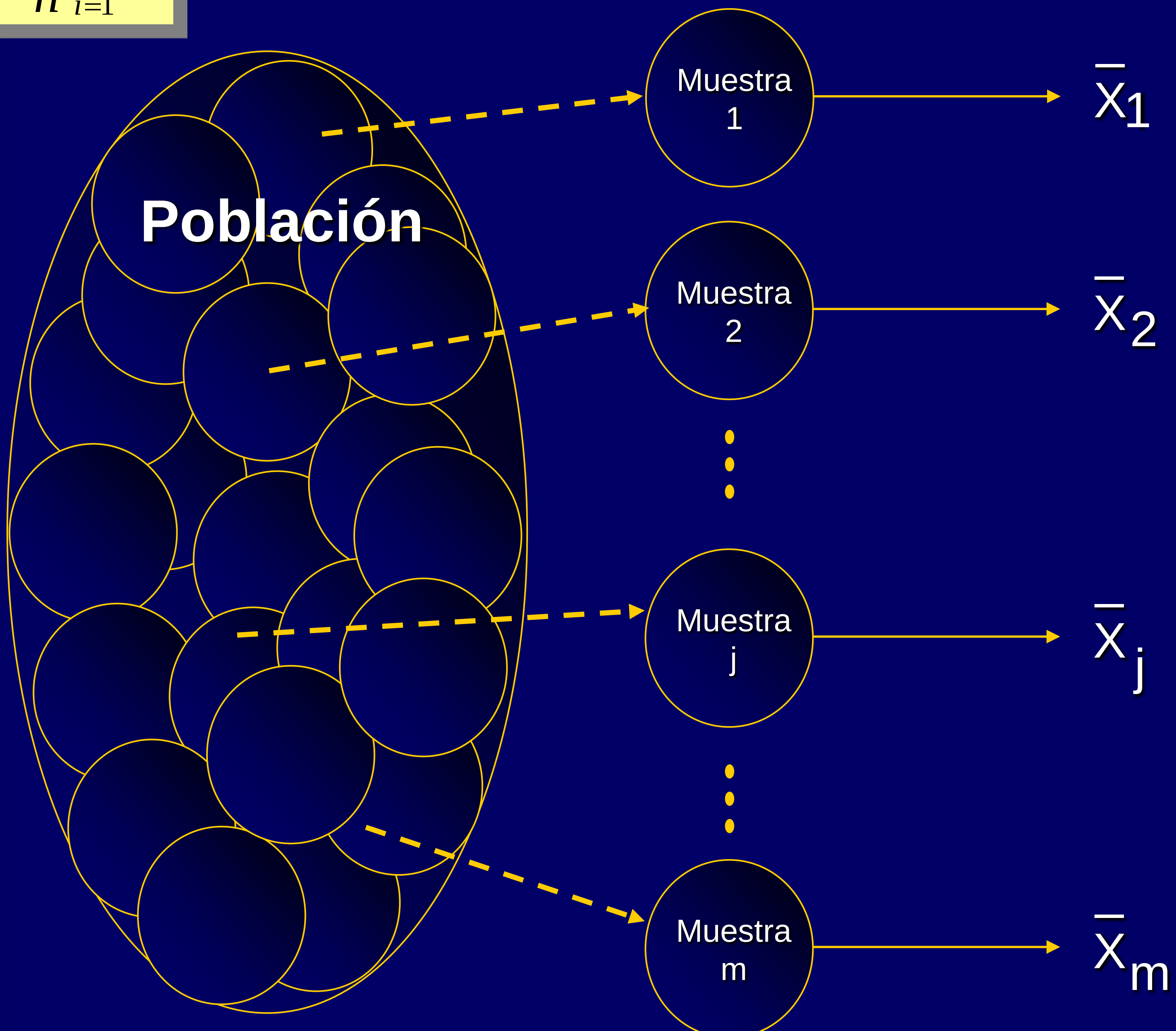
calificación
4
5
6
7
8

Obtener muestras aleatorias (con reemplazo) de tamaño 2 y para cada una de ellas calcular el valor del estadístico:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Cálculo del valor del estadístico $\overline{X}$



- Estadístico

  En consecuencia, un estadístico es una VARIABLE ALEATORIA y por lo tanto, tiene una distribución de probabilidad.

Si la población es discreta y finita, entonces la distribución de muestreo también es discreta, y existe un número finito de posibles muestras.

Si la población es continua, entonces la distribución de muestreo también es continua, ya que existe un número infinito de posibles muestras, y a su vez, el estadístico puede tomar un número infinito de valores.

Media, Varianza y Error estándar de un estadístico

Un estadístico, al ser una variable aleatoria tiene una distribución de probabilidad.

La media del estadístico es su valor esperado y su varianza es el segundo momento de la variable aleatoria con respecto a su media.

El error estándar de un estadístico es la raíz de la varianza, y es una medida de la variabilidad del estadístico.

- <u>Distribución de muestreo</u>

  Es la distribución de probabilidad de una estadístico.

E J E M P L O

Sea una población de 5 calificaciones independientes

calificación
4
5
6
7
8

Construir la distribución de muestreo del estadístico:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Encontrar el valor esperado de la media muestral: $\mathrm{E}[\overline{\mathrm{X}}]$

Encontrar la varianza de la media muestral: $\mathrm{V}[\overline{\mathrm{X}}]$

Encontrar el error estándar de la media muestral: $\mathrm{ES}[\overline{\mathrm{X}}]=\mathrm{Raiz}(\mathrm{V}[\overline{\mathrm{X}}])$

Se trata de una población discreta y fínita (consta de 5 elementos). El número de posibles muestras con reemplazo es: $m=5^2=25$

- Determinar todas las posibles muestras.
- Para cada una de ellas calcular el valor del estadístico correspondiente.
- Para cada valor del estadístico, determinar la probabilidad de que ocurra ese valor al extraer una muestra.
- Utilizando la distribución de muestreo se encuentra el valor esperado y la varianza del estadístico.

Una vez que se conoce la forma de la distribución de muestreo de un estadístico, ya se puede utilizar para realizar inferencias por intervalos y pruebas de hipótesis.

Si de una población con distribución $f(x)$ con media μ y varianza σ^2 se extrae una muestra aleatoria simple de tamaño n, entonces:

$$E[\bar{X}] = \mu$$

y

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

Demostrar . . .

EJEMPLO

Si de una población con distribución normal, con media μ y varianza σ^2, se extrae una muestra aleatoria simple de tamaño n, determinar la distribución de muestreo del estadístico:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Estimador

Un estimador es un estadístico que se utiliza para inferir el valor de algún parámetro poblacional, y que proporciona un único valor como estimación de ese parámetro.

Un ejemplo de estimador puntual es la media muestral.

Para que un estadístico funcione como estimador de algún parámetro debe al menos ser evaluado. Lo deseable es que su valor esperado sea igual al parámetro que pretende estimar y que su varianza sea mínima.

Estimador

Se dice que un estimador es insesgado si su valor esperado es igual al parámetro que pretende estimar.

Esto significa que si realizamos un número grande de experimentos (extracciones de la población); en promedio, el valor proporcionado por el estimador será muy cercano al verdadero valor del parámetro, con un error en la estimación medido por el *error estándar* del estimador.

En resumen . . .

De los ejemplos anteriores, se desprende que la media muestral $\overline{X}$ es un buen estimador de la media poblacional μ. ya que:

$$E[\overline{X}] = \mu$$

y su error estándar está dado por:

$$ES[\overline{X}] = \sqrt{\frac{\sigma^2}{n}}$$

Además, si la población de origen tiene distribución normal, entonces, por la propiedad reproductiva de la distribución normal, la distribución de probabilidad de $\overline{X}$ también tiene distribución normal.

¿ y si la población no tiene distribución normal ?

Teorema central del límite:

Sea $\underline{X}=\{x_1, x_2, . . ., x_n\}$ una muestra aleatoria simple de una población con función de densidad de probabilidad $f(x)$ cualquiera con media μ y varianza σ^2. El estadístico $\overline{X}$ tiene media μ y varianza σ^2/n y su distribución de probabilidad tiende a la de una distribución normal conforme n tiende a infinito.