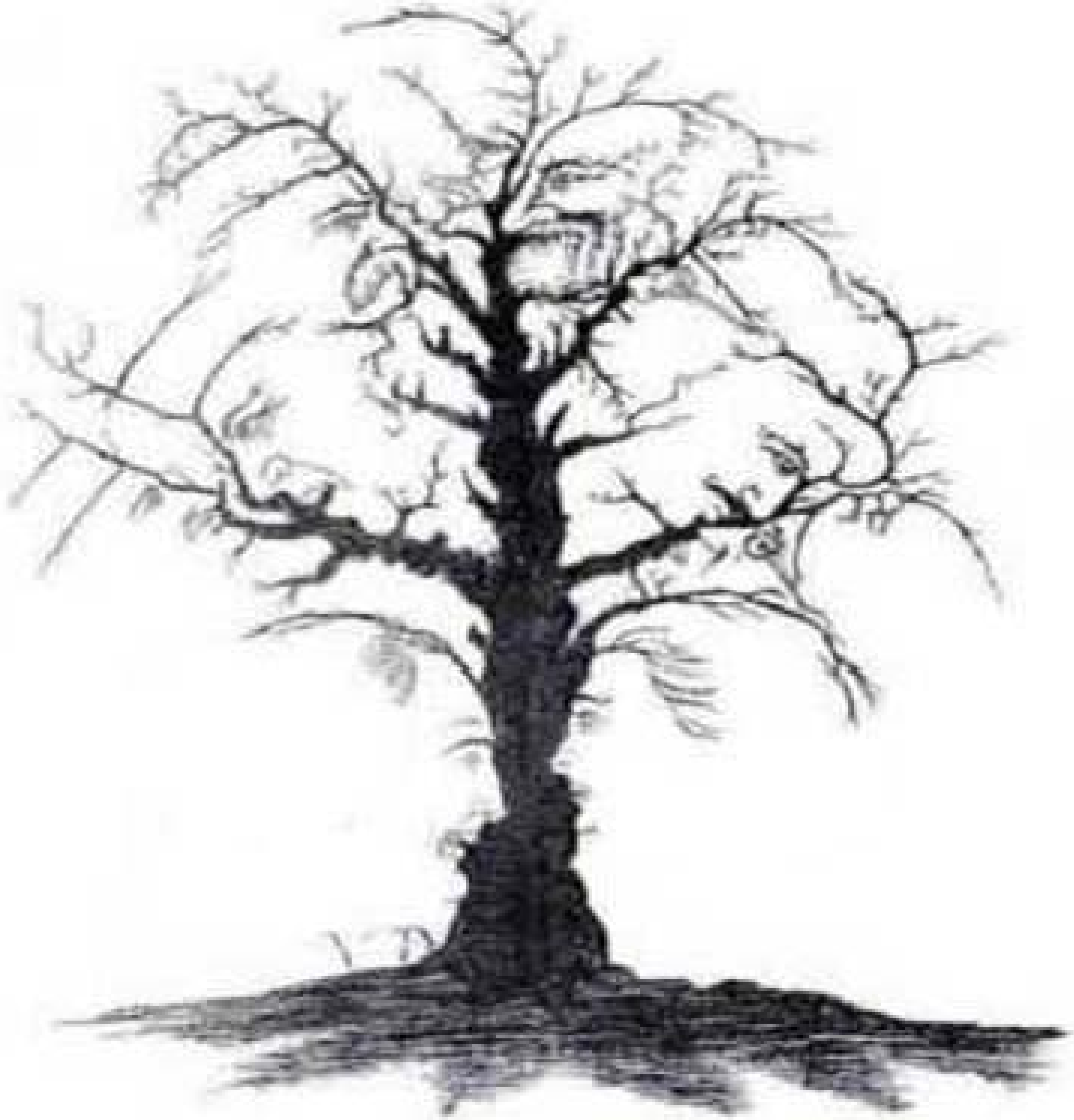


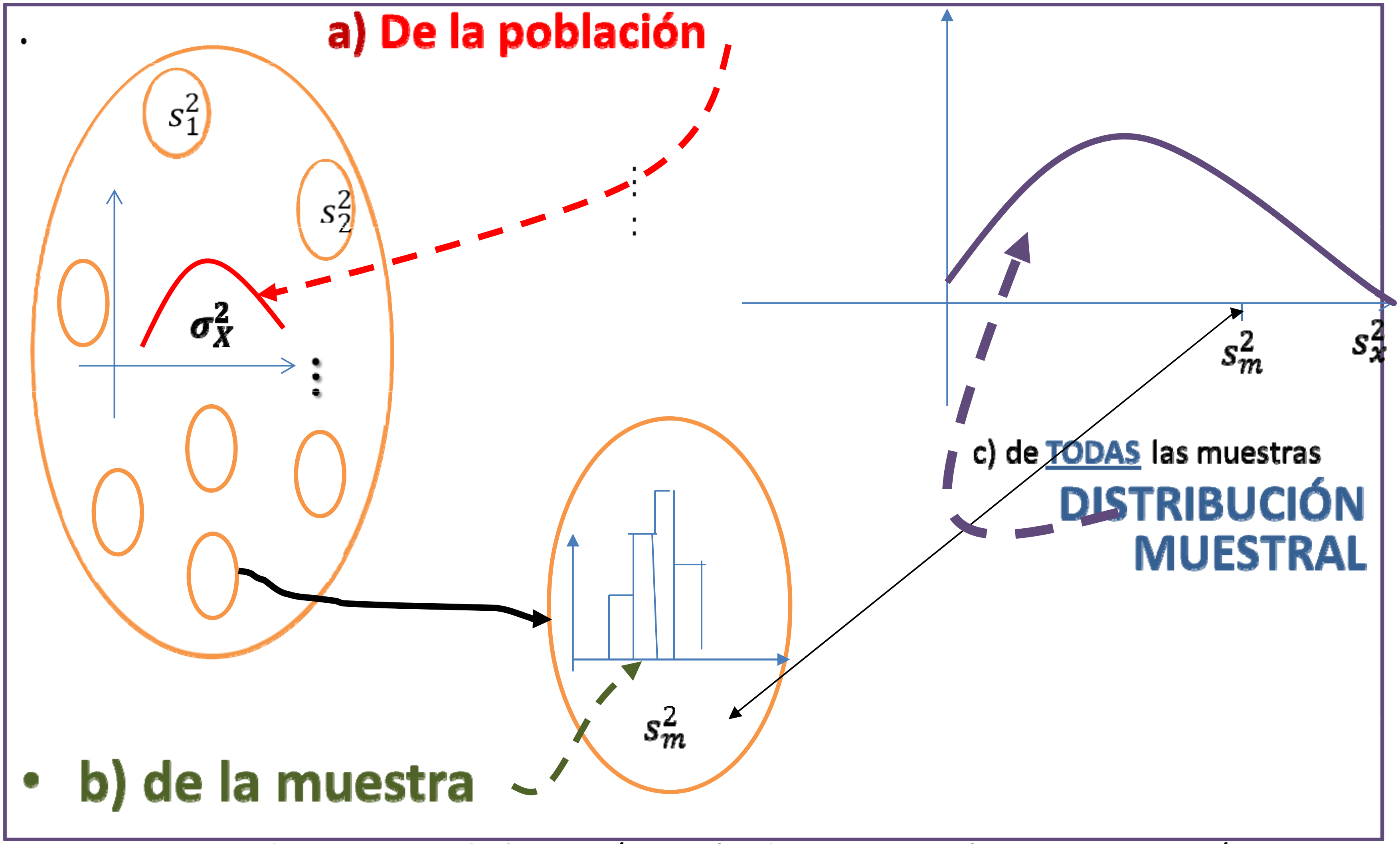
CONFERENCIA-TALLER
LAS DISTRIBUCIONES MUESTRALES I
LAS DISTRIBUCIONES MUESTRALES DE LA MEDIA Y DE LA VARIANZA

Bernardo Frontana de la Cruz
División de Ciencias Básicas, FI-UNAM
130503

Clasificación de la ESTADÍSTICA

- *Estadística*
  - *descriptiva*
    - *paramétrica*
      - *univariable*
      - *multivariable*
    - *no paramétrica*
      - *univariabl*
      - *multivariab*
  - ***inferencial***
    - ***paramétrica***
      - ***univariabl***
      - ***multivariab***
    - *no paramétrica*
      - *univariabl*
      - *multivariab*

LAS TRES DISTRIBUCIONES
de la investigación estadística



MUESTREO

- Se efectúa para:
  - Evitar el sesgo
  - Ahorrar recursos humanos, tecnológicos y financieros.

- TIPOS DE MUESTREO

- ALEATORIO SIMPLE
- SISTEMÁTICO
- POR CONGLOMERADOS
- ESTRATIFICADO

ESTADÍSTICO (1/2)

- ES UNA FUNCIÓN DE UNA MUESTRA ALEATORIA QUE **NO** CONTIENE PARÁMETROS

$$G = f(X_1, X_2, X_3, \dots, X_n)$$

- Ejemplo:
- $\bar{X} = \dfrac{\sum X_1}{n}$
- Un valor del estadístico es:

$$\bar{x} = \frac{\sum x_1}{n}$$

Muestra aleatoria

$[X_1, X_2, X_3, \dots, X_n]$

1

n

muestra

$[x_1, x_2, x_3, \dots, x_n]$

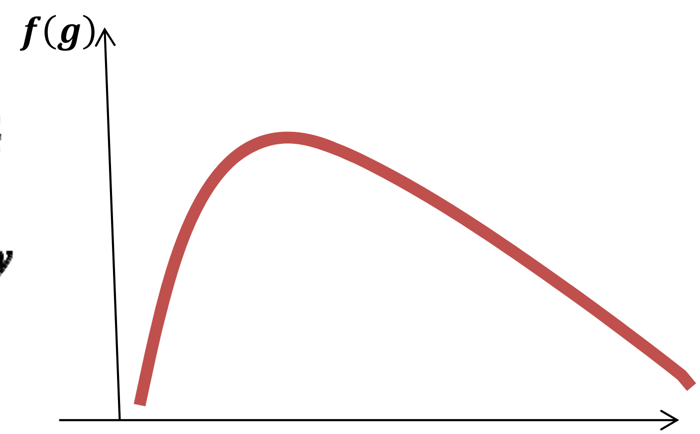
ESTADÍSTICO (2/2)

- **AL SER UNA FUNCIÓN DE UNA MUESTRA ALEATORIA, G ES UNA VARIABLE ALEATORIA:**

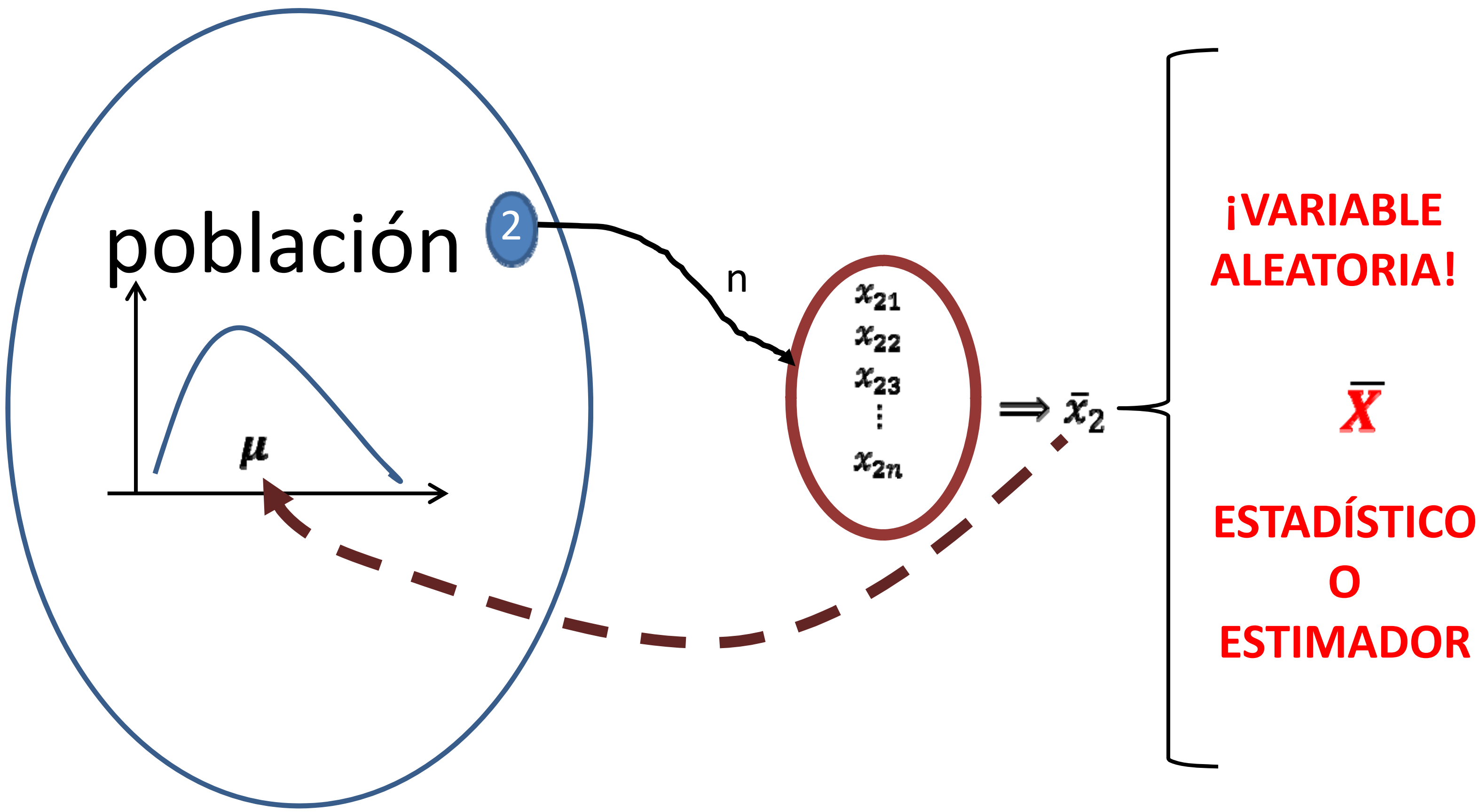
$$G = f(X_1, X_2, X_3, \dots, X_n)$$

- Tiene función de probabilidad: **DISTRIBUCIÓN MUESTRAL**
  - Distribución muestral discreta
  - Distribución muestral continua (Densidad

- $G = \begin{cases} \bar{X} \\ S_x^2 \\ \vdots \\ R_{xy} \end{cases}$ *al sacar la muestra* $g = \begin{cases} \bar{x} \\ s_x^2 \\ \vdots \\ r_{xy} \end{cases}$

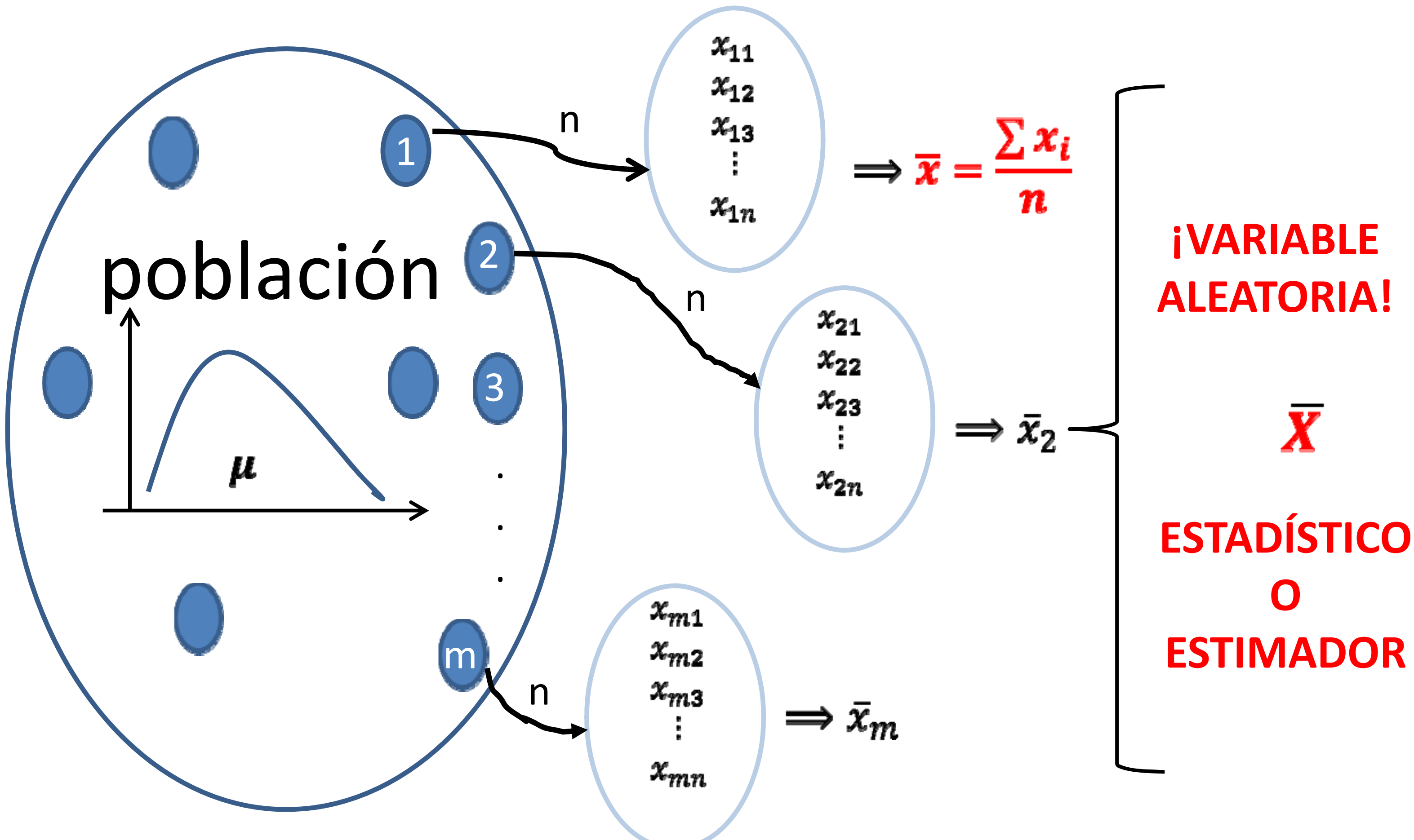


ESTADÍSTICO MEDIA O ESTIMADOR DE LA MEDIA DE LA POBLACIÓN



ESTADÍSTICO MEDIA

$$\bar{X} = \frac{\sum X_i}{n}$$



Parámetros de la Distribución Muestral de la Media

- $G = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- Aplicando el operador valor esperado:
- $E[G] = E[\bar{X}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$
- Como $E[X_i] = \mu$; se tiene $E[\bar{X}] = \mu_{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$

Si ahora aplicamos el operador varianza:

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

- Pero como cada X_i su distribución es la de la población porque aún no se saca la muestra, por lo tanto
  - $V[X_i] = \sigma^2$
- Para toda X_i se tiene
  - $V[\bar{X}] = \sigma_{\bar{X}}^2 = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

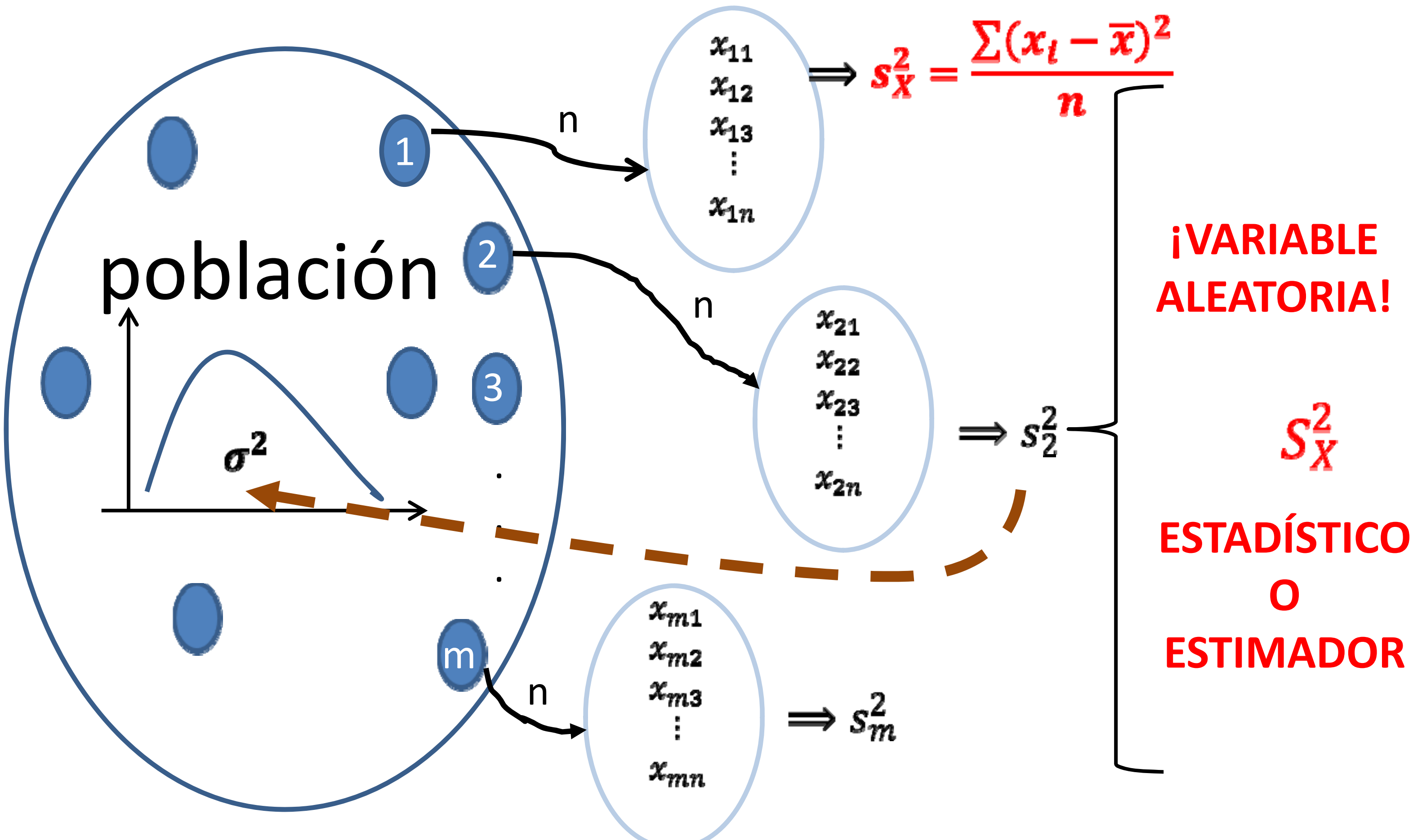
Distribución muestral de la media

- Por el Teorema central del límite:

- $\bar{X} \sim N\left(\mu_{\bar{X}} = \mu_X; \sigma_{\bar{X}}^2 = \dfrac{\sigma_X^2}{n}\right)$

ESTADÍSTICO VARIANZA O ESTIMADOR DE LA VARIANZA DE LA POBLACIÓN

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



DISTRIBUCIÓN Ji-Cuadrada

- **La Distribución χ^2_υ, Ji-Cuadrada o Chi-Cuadrada**

- La distribución normal es una distribución madre porque a parir de ella nacen otras distribuciones como es el caso de la distribución Ji-Cuadrada de amplia utilización porque modelan fielmente el comportamiento de varios estadísticos.

- Sea X_1 una VA que se distribuye $X_1 \simeq N(\mu, \sigma)$, si se estandariza se tiene $Z_1 \simeq N(0,1)$ y si se eleva al cuadrado se obtiene una nueva variable Z_1^2que **no** se distribuye normalmente y se llama $\boldsymbol{\chi^2_{\upsilon=1}}$; continuemos esta operación hasta sumar n variables.

$$X_1 \simeq N(\mu, \sigma) \Longrightarrow Z_1 \simeq N(0,1) \Longrightarrow Z_1^2 = \boldsymbol{\chi^2_{\upsilon=1}}$$
$$X_2 \simeq N(\mu, \sigma) \Longrightarrow Z_2 \simeq N(0,1) \Longrightarrow Z_2^2 + Z_1^2 = \boldsymbol{\chi^2_{\upsilon=2}}$$
$$X_3 \simeq N(\mu, \sigma) \Longrightarrow Z_3 \simeq N(0,1) \Longrightarrow Z_3^2 + Z_2^2 + Z_1^2 = \boldsymbol{\chi^2_{\upsilon=3}}$$

……………………………………………….

- $X_n \simeq N(\mu, \sigma) \Longrightarrow Z_n \simeq N(0,1) \Longrightarrow Z_n^2 + \cdots + Z_3^2 + Z_2^2 + Z_1^2 = \boldsymbol{\chi^2_{\upsilon=n}} = \sum_{i=1}^{n} Z_n^2$

  - O sea que la variable aleatoria $\chi^2_{\upsilon=n}$ se define como la suma de n variables aleatorias estandarizadas al cuadrado con $\boldsymbol{\upsilon = n}$ **grados de libertad**.

$$f(\chi^2_\upsilon) = \frac{1}{\upsilon}(\chi^2_\upsilon)^{\left(\frac{1}{2}\upsilon - 1\right)} e^{-\frac{\chi^2_\upsilon}{2}} \quad para\ \chi^2_\upsilon \geq 0$$

Distribución Muestral de la Varianza (1/2)

Para derivar la distribución muestral del estadístico varianza suponiendo que la población de la cual sacamos las muestras tiene una distribución normal y que conocemos el valor de la media μ. Considerando una sola desviación de μ tenemos

- $(X_i - \mu) = (X_i - \bar{X}) + (\bar{X} - \mu)$
- Elevando al cuadrado esta desviación y desarrollándolo
- $(X_i - \mu)^2 = [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 = (X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - \mu) + (\bar{X} - \mu)^2$
- Sumado todas las desviaciones de la media y aplicando el operador sumatoria $\sum(X_i - \mu)^2 = \sum(X_i - \bar{X})^2 + 2\sum(X_i - \bar{X})(\bar{X} - \mu) + \sum(\bar{X} - \mu)^2$
- Como $(\bar{X} - \mu)$ es constante,
- $\sum(X_i - \mu)^2 = \sum(X_i - \bar{X})^2 + 2(\bar{X} - \mu)\sum(X_i - \bar{X}) + n(\bar{X} - \mu)^2$
- Como $2(\bar{X} - \mu)\sum(X_i - \bar{X}) = 2(\bar{X} - \mu)(\sum X_i - \sum \bar{X}) = 2(\bar{X} - \mu)(n\bar{X} - n\overline{X}) = 0$
- Se tiene $\sum(X_i - \mu)^2 = \sum(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$
- Su ahora dividimos entre σ^2, $\frac{\sum(X_i-\mu)^2}{\sigma^2} = \frac{\sum(X_i-\bar{X})^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2}$

Distribución Muestral de la Varianza (2/2)

- Para el término del lado izquierdo de la expresión anterior, $\frac{\sum(X_i-\mu)^2}{\sigma^2} = \sum(\frac{X_i-\mu}{\sigma})^2 = \sum Z_i^2 = \chi^2_{v=n}$

Para el miembro derecho de la parte derecha se tiene $\frac{n(\bar{X}-\mu)^2}{\sigma^2} = \frac{(\bar{X}-\mu)^2}{\sigma^2/n} = (\frac{\bar{X}-\mu}{\sigma/\sqrt{n}})^2 = (Z)^2 = \chi^2_{v=1}$

- Sustituyendo las expresiones se tiene $\chi^2_{v=n} = \frac{\sum(X_i-\bar{X})^2}{\sigma^2} + \chi^2_{v=1}$
- Por la propiedad de aditividad se tiene $\frac{\sum(X_i-\bar{X})^2}{\sigma^2} = \chi^2_{v=n-1}$
- Finalmente, si de la definición de varianza despejamos $\sum_{i=1}^{n}(X_i-\bar{X})^2$ y la sustituimos en la anterior se obtiene que $\frac{nS^2}{\sigma^2} = \chi^2_{v=n-1}$
- ***Por lo que hemos demostrado que si $\bar{X}$ y S^2 devienen de una población que se distribuye normalmente, entonces la relación $\frac{nS^2}{\sigma^2}$ se distribuye conforme una distribución Chi-cuadrada con $v = n-1$ grados de libertad.***
- Si en lugar de utilizar S^2 utilizamos el estimador insesgado $\hat{S}^2$ se obtiene $\frac{(n-1)\hat{S}^2}{\sigma^2} = \chi^2_{v=n-1}$
- Usualmente en los trabajos estadísticos sobre inferencias se utilizan los cocientes de estas expresiones porque son variables Chi-cuadradas que se encuentran directamente en las tablas o en la computadora.

Ejemplo de distribución muestral de la media

- Si la edad de los académicos de la Facultad de Ingeniería tiene una media de 45 años y una desviación estándar de 15, para una muestra aleatoria de 25 académicos, se tiene que $E[\bar{X}] = \mu = 45$ y $V[\bar{X}] = \frac{\sigma^2}{n} = \frac{15^2}{25} = 9$. Al sacar la muestra se tiene $\bar{x} = 48$ años, cuyo valor estandarizado es
  - $Z = \frac{48-45}{\frac{15}{\sqrt{25}}} = 1.$
- Si aumentamos la muestra a 625 académicos y resulta también $\bar{x} = 48$, el nuevo valor estandarizado será
  - $Z = \frac{48-45}{\frac{15}{\sqrt{625}}} = 5.$
- En ambos casos la media de la muestra está 3 años alejada de la media de la población -$48 - 45 = 3$- pero al aumentar la muestra el valor estandarizado esta mucho más alejado de 0, lo que sorprende observar esta diferencia cuando la muestra es grande. Como hasta ahora solo conocemos la media y la varianza de la distribución muestral de la media, con la desigualdad de Chebyshev podemos estimar la probabilidad de este evento.
-
- $p(|\bar{X} - \mu| < 3) \geq 1 - \frac{\sigma_{\bar{X}}^2}{9} \geq 1 - \frac{0.360}{9} \geq 0.96$
-
- O sea que la probabilidad de que la diferencia $|\bar{X} - \mu| < 3$ sea ≥ 0.96 , así, $\bar{x} = 48$ años es un evento inusual. Este tipo de razonamientos son muy útiles cuando se hacen inferencias de la media de la población.

Ejemplo de la distribución muestral de la varianza

- Se utilizan medidas antropométricas para determinar el alcance entre los tableros de control y los operadores, si se toma una muestra de $n = 25$ operadores sentados de una población normal y se les mide la distancia efectiva, aunque la varianza de la población se desconoce, los registros de las investigaciones previas sugieren que la varianza de la población es de $\sigma^2 = (4\ cm)^2 = 16\ cm^2$. La probabilidad de que la varianza de la muestra exceda 20 cm^2, usando la expresión es
-
- $\alpha = p(\hat{S}^2 > 20 | \sigma^2 = 16\ cm^2) = \text{p}[\frac{(n-1)\hat{S}^2}{\sigma^2} > \frac{20\ (25-1)}{16}] = p(\chi^2_{v=24} > 30)$
-
- De la tablas de la distribución Chi-cuadrada, para $n = 25$, se tienen $v = 25 - 1 = 24$ g. de l. con lo cual la probabilidad está dentro del intervalo
-
- $0.10 < p(\hat{S}^2 > 20) < 0.20$