Universidad Nacional de La Plata

Facultad de Ciencias Agrarias y Forestales

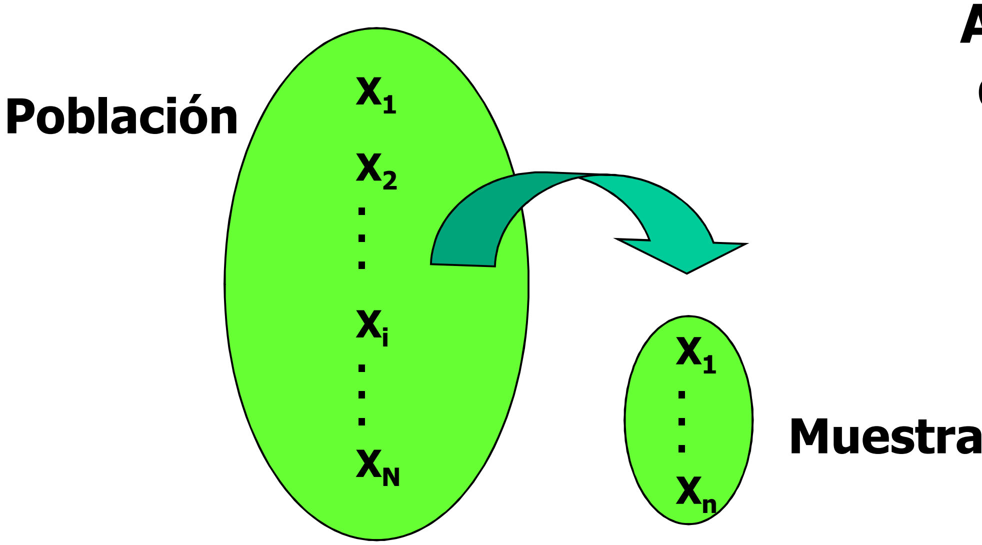
CÁLCULO ESTADÍSTICO Y BIOMETRÍA

ESTADÍSTICA

DESCRIPTIVA

INFERENCIAL

Auxilio de la Teoría de Probabilidades



Diseño del Muestreo

Como se seleccionarán las muestras ??	Cuantas unidades formarán la muestra ??
Muestreo aleatorio simple	Variabilidad de los datos
Muestreo aleatorio estratificado	Error admisible
Muestreo por conglomerados	Confianza
Muestreo sistemático	

INFERENCIA ESTADISTICA

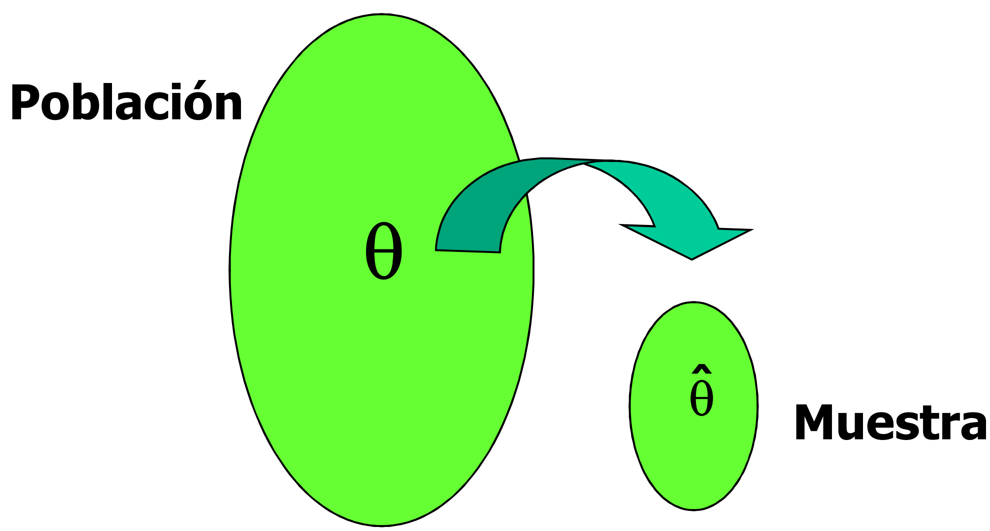
Estimación de Parámetros

- Propiedades de un estimador
- Estimación a Intervalo de la media poblacional
- Estimación a Intervalo de la varianza poblacional

Pruebas de Hipótesis

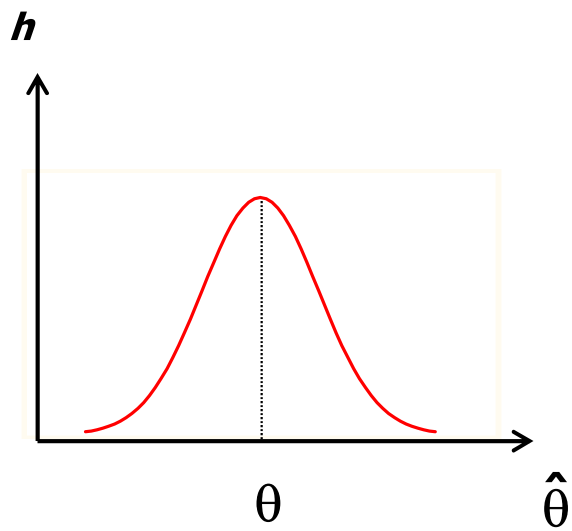
Estimación de Parámetros

Estimador: es un estadístico o función obtenido de los datos de una muestra y que pretendemos que represente lo más fielmente posible a un parámetro poblacional desconocido, que es el objeto de nuestro estudio.



Propiedades de un buen estimador:

Ser INSESGADO: un estimador es insesgado cuando su valor esperado es igual al parámetro que estima, es decir: $E(\hat{\theta}) = \theta$



$$E(\bar{x}) = \mu \qquad E(s^2) \neq \sigma^2$$

con $$s^2_x = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

$$E(s^2) = \frac{(n-1) \cdot \sigma^2}{n}$$

En cambio si : $$s^2_x = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$ entonces: $$E(s^2) = \sigma^2$$

- **Ser de MINIMA VARIANCIA (eficiente):**

si $\hat{\theta}_1$ **y** $\hat{\theta}_2$ **son 2 estimadores insesgados de** θ **se dice que** $\hat{\theta}_1$ **es mejor que** $\hat{\theta}_2$ **si** $V(\hat{\theta}_1) < V(\hat{\theta}_2)$

- **Ser CONSISTENTE:**

$\hat{\theta}$ **será mejor a medida que aumenta el tamaño de muestra, esto es a medida que** n **aumenta la distribución de** $\hat{\theta}$ **se concentra más alrededor de** θ

$$\lim_{n \to \infty} \hat{\theta} = \theta \qquad\qquad \lim_{n \to \infty} P\left(\left| \hat{\theta} - \theta \right| = 0 \right) = 1$$

Cuando se propone algún estimador se debe estudiar su distribución para ver como se comporta

Intervalos de confianza

para la media poblacional (μ)

Distribución de la media muestral

$\bar{x}$ es una V.A. con parámetros distribucionales:

$$E(\bar{x}) = \mu_{\bar{x}} = \mu_x \qquad V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Teorema Central del Límite : Si $x_1, x_2, \ldots, x_n$ es una muestra aleatoria de tamaño n de una población con media μ y variancia σ^2 y si $\bar{x}$ es la media muestral, entonces la forma límite de la distribución de Z definida como:

$$\boxed{Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$$

cuando $n \to \infty$ la variable Z tiende a la distribución normal estándar N(0;1)

1a. Intervalo de confianza para μ con σ^2 conocida

Conociendo $\bar{x}$ establecer x_1 y x_2 tal que:

$$\boxed{P(x_1 < \mu_x < x_2) = \gamma = 1 - \alpha}$$

Operando →

$$\boxed{\ell_{i(\mu)} = \bar{x} \pm z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}}$$

Tamaño de muestra óptimo:

$$\boxed{n = \left(\frac{z.\sigma}{EA} \right)^2}$$

EA: Error admisible o de estimación

Ejercicio 1

- El diámetro de las tortas de girasol se distribuye normalmente con desvío estándar de 6 cm. Sobre una muestra de 10 tortas, se obtuvo un promedio de diámetro de 18 cm encontrar:

a) El intervalo de confianza (IC) del 95% para la media poblacional

b) El tamaño de muestra para que el IC informe con un error de estimación no mayor a 2 cm

1b. Intervalo de confianza para μ con σ^2 desconocida

Conociendo $\bar{x}$ y S establecer x_1 y x_2 tal que:

$$\boxed{P(x_1 < \mu_x < x_2) = \gamma = 1 - \alpha}$$

Al sustituir σ por S se genera una nueva variable aleatoria denominada t definida como:

$$\boxed{t_0 = \frac{\bar{x} - \mu}{S / \sqrt{n}}}$$

Con $t_0 \sim t\,(n-1)$

Esta nueva variable ya no tiene distribución normal sino una distribución conocida como **t de Student** con n – 1 grado de libertad

Operando →

$$\boxed{\ell_{i(\mu)} = \bar{x} \pm t_{(1-\alpha/2;n-1)} \frac{S}{\sqrt{n}}}$$

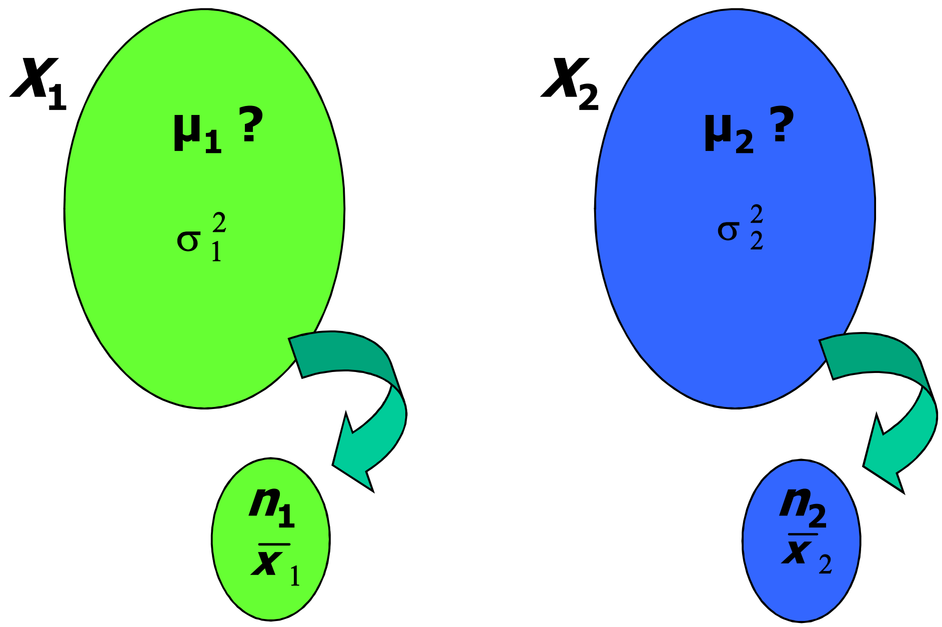
Ejercicio 2

- Se desea establecer el contenido vitamínico promedio de un alimento balanceado para pollos. Para ello se toma una muestra de 41 bolsas y se encuentra que el contenido promedio de vitaminas cada 100 gr. es = 12 mg. y que la desviación estándar S=2 mg.

a) Construir un IC del 95%.

b) Obtener un IC del 99%.

2. Intervalo de confianza para diferencias entre dos μ



$$\boxed{Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}}$$

$$\boxed{t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}}$$

2a. Intervalo de confianza para diferencias entre dos μ con varianzas poblacionales conocidas (iguales o distintas)

$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Con $Z_0 \sim N(0;1)$

$$\ell_{i(\mu_1-\mu_2)} = (\bar{x}_1 - \bar{x}_2) \pm z_{(1-\alpha/2)} \sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}$$

2b. Intervalo de confianza para diferencias entre dos μ con varianzas poblacionales desconocidas pero iguales

Considerando $n_1 = n_2$ o $n_1 \neq n_2$

En este caso la variable aleatoria t queda definida como:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Con $t_0 \sim t\,(n_1 + n_2 - 2)$

$$\ell_{i(\mu_1-\mu_2)} = (\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2\,;\,n_1+n_2-2)} \sqrt{S_p^2 \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

donde:

$$s_p^2 = \frac{(n_1 - 1)\,s_1^2 + (n_2 - 1)\,s_2^2}{n_1 + n_2 - 2}$$

Ejercicio 3

- Se realiza un ensayo en maíz en el cual se aplica fertilización a 15 parcelas experimentales y otras 17 parcelas no se fertilizan. Al finalizar el ensayo se registran los valores de la variable en estudio, en mg.

	$\bar{X}$	S^2
Con Fertilización	311	1953,95
Sin Fertilización	261,98	1722,82

a) Considerando varianzas poblacionales iguales, construir un IC para la diferencia de medias del 95%.

2c. Intervalo de confianza para diferencias entre dos μ con varianzas poblacionales desconocidas pero iguales

Considerando $n_1 = n_2$ podemos simplificar los cálculos a:

En este caso la variable aleatoria t queda definida como:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{2.S^2}{n}}}$$

Con $t_0 \sim t\,(n_1+n_2-2)$

$$\ell_{i(\mu_1-\mu_2)} = (\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2\,;n_1+n_2-2)} \sqrt{\frac{2.S^2}{n}}$$

donde:

$$S^2 = \frac{s_1^2 + s_2^2}{2}$$

2d. Intervalo de confianza para diferencias entre dos μ con varianzas poblacionales desconocidas y distintas

En este caso la variable aleatoria t queda definida como:

$$t_0^* = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

Con $t^*_0 \sim t\,(v)$

Se distribuye aproximadamente como t con ν grados de libertad, donde :

$$\nu = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_1^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}} - 2$$

donde:

$$\ell_{i(\mu_1-\mu_2)} = (\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2\,;\nu)} \sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}$$

Intervalo de confianza para la varianza poblacional (σ^2)

Distribución de la varianza muestral

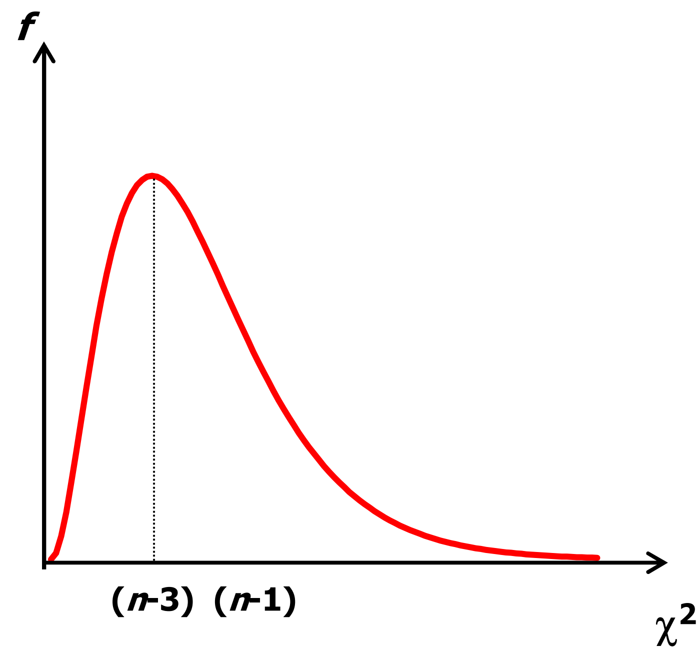
S^2 es una V.A. con parámetros distribucionales:

$$E(S^2) = \mu_{S^2} = \sigma^2$$

Al igual que en el caso de las medias muestrales, la idea es visualizar la distribución de las varianzas muestrales y poder identificar un modelo de probabilidad al cual se ajuste la distribución. Dicho ajuste no se realiza sobre la distribución de S^2 sino sobre el estadístico Chi-cuadrado χ^2

Empírica: $\boxed{\dfrac{(n-1)\cdot s^2}{\sigma^2}} \sim \chi^2(n-1)$

- $E(\chi^2) = n - 1$
- $Mo(\chi^2) = n - 3$
- $V(\chi^2) = 2(n-1)$

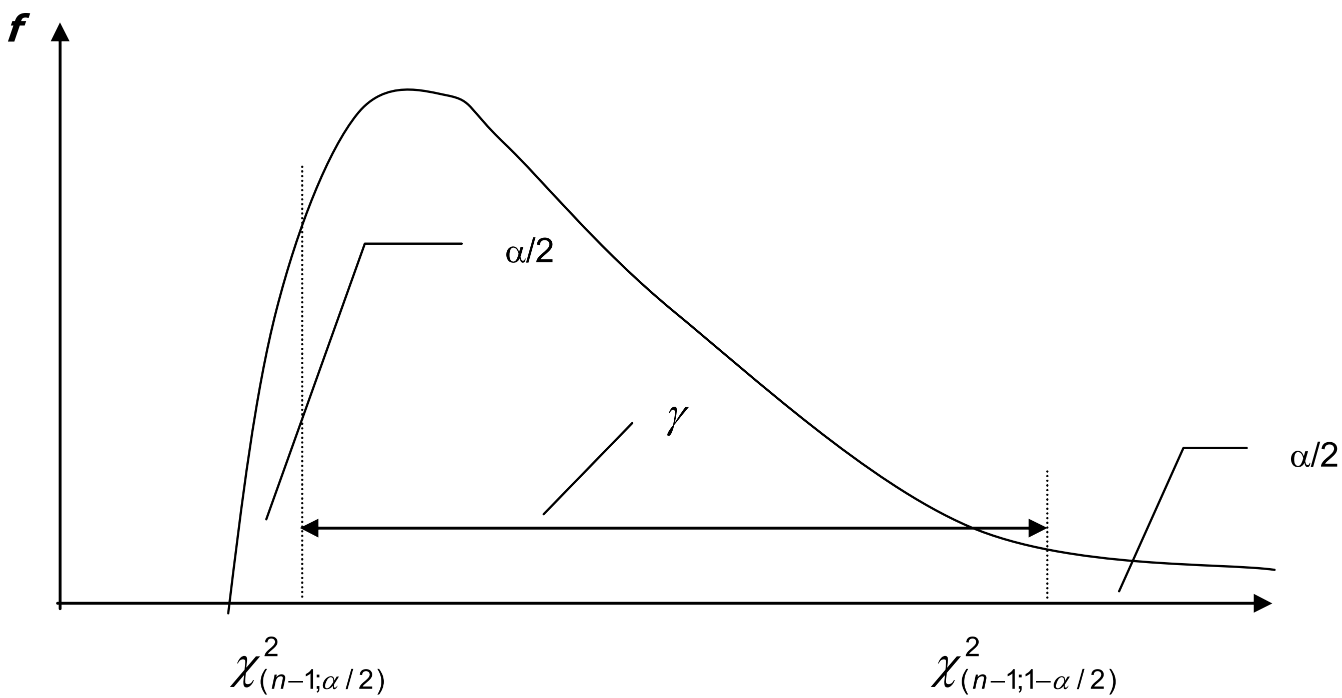


Intervalo de confianza para la varianza muestral

Conociendo S^2 establecer x_1 y x_2 tal que:

$$\boxed{P(x_1 < \sigma^2 < x_2) = \gamma = 1 - \alpha}$$

$$\boxed{P\left[\frac{(n-1)\cdot S^2}{\chi^2_{(n-1;1-\alpha/2)}} \leq \sigma^2 \leq \frac{(n-1)\cdot S^2}{\chi^2_{(n-1;\alpha/2)}}\right] = \gamma}$$



Continuación del Ejercicio 2

- Se desea establecer el contenido vitamínico promedio de un alimento balanceado para pollos. Para ello se toma una muestra de 41 bolsas y se encuentra que el contenido promedio de vitaminas cada 100 gr. es = 12 mg. y que la desviación estándar S=2 mg.

a) Construir un IC del 95% para la varianza poblacional
b) Obtener un IC del 95% para el desvío estándar poblacional