TEMA 7. ESTIMACIÓN

❖ 7.1. Introducción y definiciones

▪ Supongamos que conocemos la distribución de la característica de interés de una población

▪ La función de densidad o masa de probabilidad depende del vector de parámetros θ : $f(x\ ;\ \theta)$

▪ Se desean estimar los parámetros a partir de una muestra

¿Cómo hacer esta estimación?

▪ **Estimación puntual:** Se busca un estimador, que con base a los datos muestrales dé origen a un valor puntual que utilizamos como estimación del parámetro

▪ **Estimación por intervalos:** Se determina un intervalo aleatorio que, de forma probable, contiene el verdadero valor del parámetro. Este intervalo recibe el nombre de **intervalo de confianza**

❑ **Ejemplo**

Sea X una v.a. que estudia el grosor del tronco de un arbusto. Se conoce que dicha variable es normal con desviación típica 1 pero no se conoce la media.

$$X \longrightarrow \mathcal{N}(\mu, 1) \quad ; \quad \mu \in \boldsymbol{R}$$

El parámetro a estimar es μ

❖ 7.2. Estimación puntual. Propiedades deseables de los estimadores

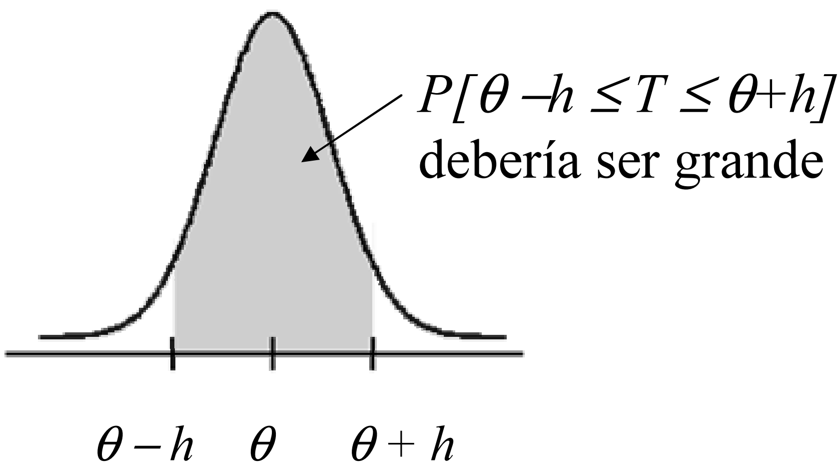
◈ 7.2.1. Introducción y definiciones

Sea $X_1, \ldots, X_n$ una muestra aleatoria simple con función de densidad $f(x;\ \theta)$

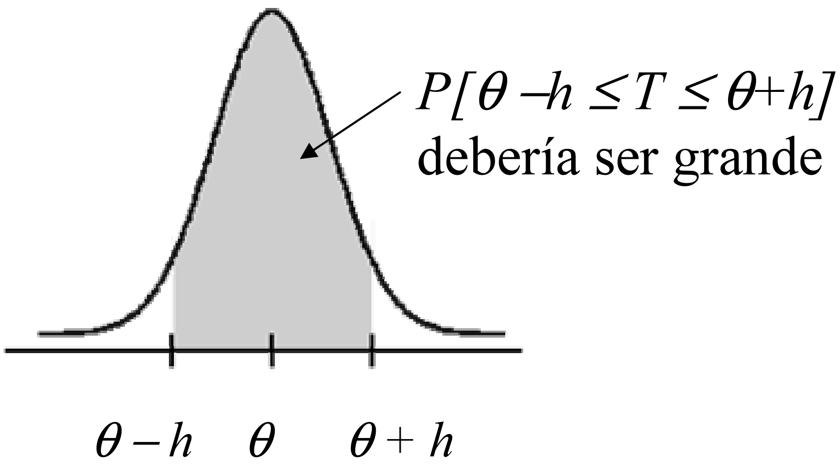
Sea un estadístico $T = u(X_1, \ldots, X_n)$

El problema es encontrar una función u que proporcione el mejor estimador de θ

El estimador, T, del parámetro θ debe tener una distribución concentrada alrededor de θ y la varianza debe ser lo menor posible



❖ Error cuadrático medio



Para estudiar la variabilidad de los valores del estimador alrededor del parámetro se hace uso de una cantidad llamada **error cuadrático medio**

Definición. *Error cuadrático medio*

T estimador de θ

$$ECM(T) = E[(T-\theta)^2] = Var[T] + [\theta - E[T]]^2$$

❑ Ejemplo

Sea $X_1, \ldots, X_n$ una muestra aleatoria simple (m.a.s.) de una población X de la que se sabe que $E[X] = \mu$ y $Var[X] = \sigma^2$.

Sea $T = \bar{X}$ un estimador de μ.

Hallar el error cuadrático medio que se comete.

Solución.-

$$ECM(\bar{X}) = Var[\bar{X}] + \left(\mu - E[\bar{X}]\right)^2$$

Calculamos en primer lugar la esperanza del estimador

$$E[\bar{X}] = E\left[\frac{\sum_i X_i}{n}\right] = \sum_i \frac{E[X_i]}{n} = \frac{n\mu}{n} = \mu$$

Por lo tanto el ECM es

$$ECM(\bar{X}) = Var[\bar{X}] = Var\left[\frac{1}{n}\sum_i X_i\right] =$$

$$= \frac{1}{n^2}\sum_i Var[X_i] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

◈ 7.2.2. Estimadores Insesgados

▪ **Definición**. *Sesgo de un Estimador*

Sea T el estimador del parámetro θ. Se define el sesgo del estimador como

$$\theta - E[T]$$

▪ **Definición**. *Estimador Insesgado*

T es un estimador insesgado de θ si y sólo si

$$E[T] = \theta \qquad \text{para todo } \theta$$

NOTA: En este caso $ECM(T) = Var[T]$

❑ Ejemplo

Sea $X_1, \ldots, X_n$ una muestra aleatoria simple (m.a.s.) de una población X de la que se sabe que $E[X] = \mu$ y $Var[X] = \sigma^2$.

1.- Demostrar que $\bar{X}$ es un estimador insesgado de μ

2.- Demostrar que la cuasivarianza S^2 es un estimador insesgado de σ^2

Solución

1.- Demostrado en el ejemplo anterior

2.-

$$E\left[S^2\right]=E\left[\frac{1}{n-1}\sum_i\left(X_i-\overline{X}\right)^2\right]=E\left[\frac{1}{n-1}\sum_i\left((X_i-\mu)-\left(\overline{X}-\mu\right)\right)^2\right]$$

$$=\frac{1}{n-1}E\left[\sum_i(X_i-\mu)^2+\sum_i\left(\overline{X}-\mu\right)^2-2\sum_i(X_i-\mu)\left(\overline{X}-\mu\right)\right]$$

$$=\frac{1}{n-1}\left[E\left[\sum_i(X_i-\mu)^2\right]+E\left[\sum_i\left(\bar{X}-\mu\right)^2\right]-2E\left[\left(\bar{X}-\mu\right)\sum_i(X_i-\mu)\right]\right]$$

$$=\frac{1}{n-1}\left[E\left[\sum_i(X_i-\mu)^2\right]+nE\left[\left(\bar{X}-\mu\right)^2\right]-2nE\left[\left(\bar{X}-\mu\right)^2\right]\right]$$

$$=\frac{1}{n-1}\left[\sum_i\sigma^2+n\frac{\sigma^2}{n}-2n\frac{\sigma^2}{n}\right]=$$

$$=\frac{1}{n-1}\left[n\sigma^2+\sigma^2-2\sigma^2\right]=\frac{(n-1)\sigma^2}{n-1}=\sigma^2$$

❖ 7.3. Estimación por intervalos de confianza

◈ 7.3.1. Introducción

Se desea calcular un intervalo aleatorio que contenga al verdadero valor del parámetro, θ, con una cierta probabilidad

$$h_1(T) \leq \theta \leq h_2(T)$$

- Las funciones h_1 y h_2 son funciones de un estadístico T relacionado con el parámetro a estimar en cada caso

- **Definición:** *Nivel de confianza* ($1-\alpha$)

El nivel de confianza, $1-\alpha$, es la probabilidad de que un intervalo de confianza contenga al verdadero valor del parámetro.

$$P\,[\,h_1(T) \leq \theta \leq h_2(T)\,] = 1-\alpha$$

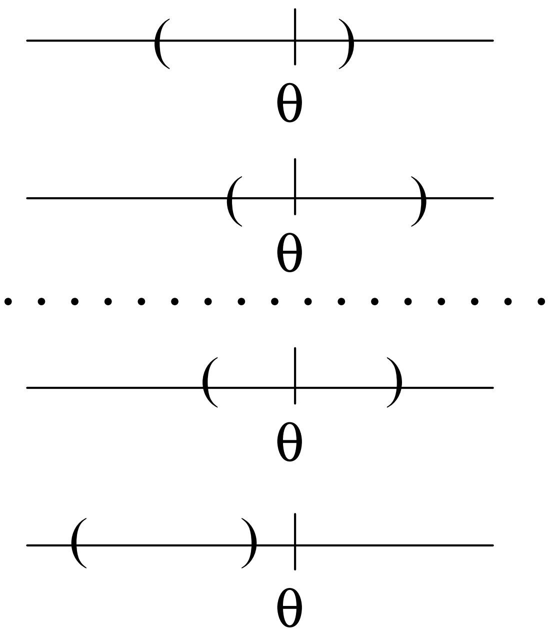
❖ **Definición:** *Intervalos de confianza bilaterales*

$$P\,[\,h_1(T) \leq \theta \leq h_2(T)\,] = 1-\alpha$$

❖ **Definición:** *Intervalos de confianza unilaterales*

$$P\,[\theta \geq h_1(T)\,] = 1-\alpha$$
$$P\,[\theta \leq h_2(T)\,] = 1-\alpha$$

➢ **NOTA:** De cada 100 intervalos construidos a partir de 100 muestras, 100 *(1–α) %* deberían contener al verdadero valor del parámetro.



Intervalo aleatorio

θ fijo

◈ 7.3.2. Intervalos de confianza para una población normal

❖ *Se muestrea una población normal para estimar los parámetros de esta población*

$X_1, \ldots, X_n$ m.a.s. de una población $X \longrightarrow N(\mu, \sigma)$

$$X_1, X_2, \ldots X_n \begin{cases} \text{Independientes entre sí} \\ X_i \longrightarrow N(\mu, \sigma) \end{cases}$$

Se desea estimar alguno de los parámetros, o ambos, según sea o no conocido el otro

➢ 7.3.2.1. Intervalos de confianza para la media

▪ Se desea estimar la media poblacional mediante un intervalo de confianza

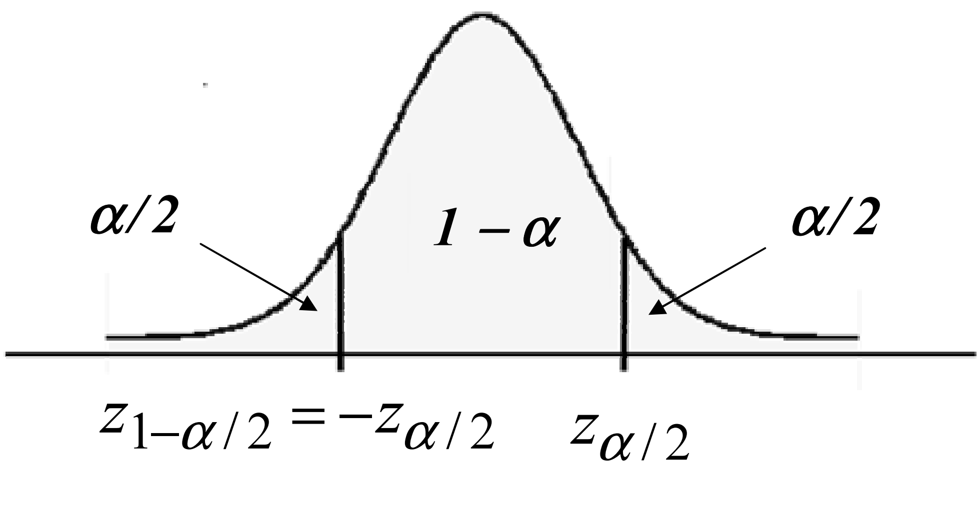
❖ Varianza poblacional conocida σ_0^2

Estadístico asociado al parámetro a estudiar μ : $T = \bar{X}$ (media muestral).

⌘ Distribución de la media muestral:

$$\bar{X} \rightarrow N\left(\mu; \frac{\sigma_0^2}{n}\right) \quad \Leftrightarrow \quad Z = \frac{\left(\bar{X} - \mu\right)}{\sigma_0 / \sqrt{n}} \rightarrow N(0;1)$$

$$P\left[-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right] = 1 - \alpha$$



$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] = P\left[-z_{\alpha/2} \le \frac{\left(\bar{X}-\mu\right)}{\sigma_0 / \sqrt{n}} \le z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[-z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \le \bar{X}-\mu \le z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] = 1-\alpha$$

$$P\left[-\bar{X}-z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \le -\mu \le -\bar{X}+z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] = 1-\alpha$$

$$P\left[\bar{X}+z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \ge \mu \ge \bar{X}-z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] = 1-\alpha$$

$$P\left[\bar{X}-z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \le \mu \le \bar{X}+z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] = 1-\alpha$$

❖ I.C. para μ, con varianza conocida, al nivel de confianza $1-\alpha$.

$$\left[\bar{X}-z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X}+z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right]$$

❖ I.C. para μ, con varianza conocida, al nivel de confianza $1-\alpha$

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right]$$

NOTA:

- A medida que aumenta el tamaño de la muestra disminuye la amplitud del intervalo

- A medida que el nivel de confianza es mayor aumenta la amplitud del intervalo

❑ **Ejemplo.** Se desea estudiar el peso en gramos del fruto producido por una planta. Para ello se tomó una muestra de 16 plantas observando los siguientes pesos: 506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496. El peso del fruto de cada planta es una v.a. Normal con desviación típica 5 gr. Obtener un intervalo de confianza al nivel de confianza 0.9 para el peso medio del fruto de esta planta.

Solución.-

I.C. para μ con varianza conocida
al nivel de confianza $1-\alpha$

$$\left[\bar{X} - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}} \right]$$

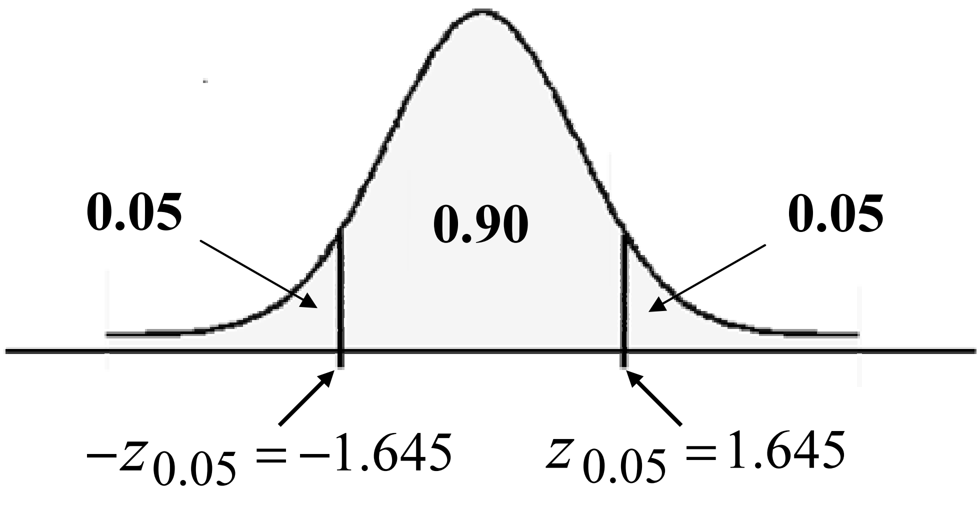
$$\bar{x} = 503.75; \quad n = 16; \quad \sigma_0{}^2 = 25; \quad \sigma_0 = 5$$

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right]$$

$$\bar{x} = 503.75; \quad n = 16; \quad {\sigma_0}^2 = 25\,; \quad \sigma_0 = 5$$

$$1-\alpha = 0.90; \quad \alpha = 0.10; \quad \alpha/2 = 0.05 \;\Rightarrow$$

$$z_{\alpha/2} = z_{0.05} = 1.645$$



$$\left[503.75 - 1.645\frac{5}{4},\ 503.75 + 1.645\frac{5}{4}\right] =$$

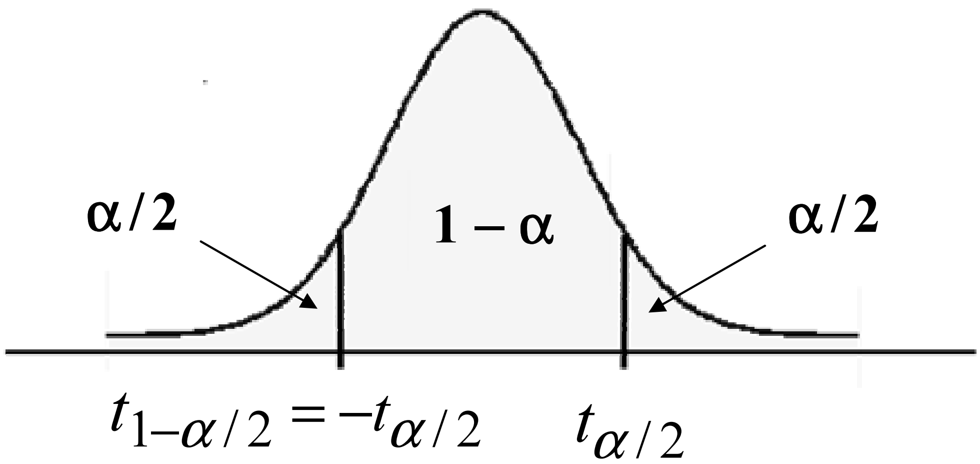
$$= \left[501.69,\ 505.81\right]$$

❖ Varianza poblacional desconocida

⌘ **Estimador:** $\overline{X}$

Distribución muestral: $T = \dfrac{\left(\bar{X}-\mu\right)}{S/\sqrt{n}} \rightarrow t_{n-1}$

Cuasidesviación típica $S = \sqrt{\dfrac{1}{n-1}\sum_i \left(X_i - \bar{X}\right)^2}$



$$P\left[-t_{\alpha/2;n-1} \le T \le t_{\alpha/2;n-1}\right] = 1-\alpha$$

$$P\left[-t_{\alpha/2;n-1} \le \frac{\left(\bar{X}-\mu\right)}{S/\sqrt{n}} \le t_{\alpha/2;n-1}\right] = 1-\alpha$$

$$P\left[-t_{\alpha/2;n-1} \leq \frac{\left(\bar{X}-\mu\right)}{S/\sqrt{n}} \leq t_{\alpha/2\,;n-1}\right]=1-\alpha$$

$$P\left[-t_{\alpha/2;n-1}\frac{S}{\sqrt{n}} \leq \bar{X}-\mu \leq t_{\alpha/2;\,n-1}\frac{S}{\sqrt{n}}\right]=1-\alpha$$

$$P\left[-\bar{X}-t_{\alpha/2;n-1}\frac{S}{\sqrt{n}} \leq -\mu \leq -\bar{X}+t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]=1-\alpha$$

$$P\left[\bar{X}+t_{\alpha/2;n-1}\frac{S}{\sqrt{n}} \geq +\mu \geq \bar{X}-t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]=1-\alpha$$

$$P\left[\bar{X}-t_{\alpha/2;n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X}+t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]=1-\alpha$$

I.C. para μ con varianza desconocida
al nivel de confianza 1–α

$$\left[\bar{X}-t_{\alpha/2;n-1}\frac{S}{\sqrt{n}},\, \bar{X}+t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]$$

❑ **Ejemplo.** Se desea estudiar el peso en gramos del fruto producido por una planta. Para ello se tomó una muestra de 16 plantas observando los siguientes pesos: 506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496. Del peso del fruto sólo se conoce que es una v.a. Normal. Obtener un intervalo de confianza al nivel de confianza 0.9 para el peso medio del fruto de esta planta.

Solución

$$\left[\overline{X} - t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]$$

I.C. para μ con varianza desconocida
al nivel de confianza 1–α

$$\overline{x} = 503.75\ ; \quad s = 6.2022\ ; \quad n = 16$$

$$\left[\overline{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}\right] =$$

$$= \left[503.75 - 1.75 \times \frac{6.2022}{4}, 503.75 + 1.75 \times \frac{6.2022}{4}\right] =$$

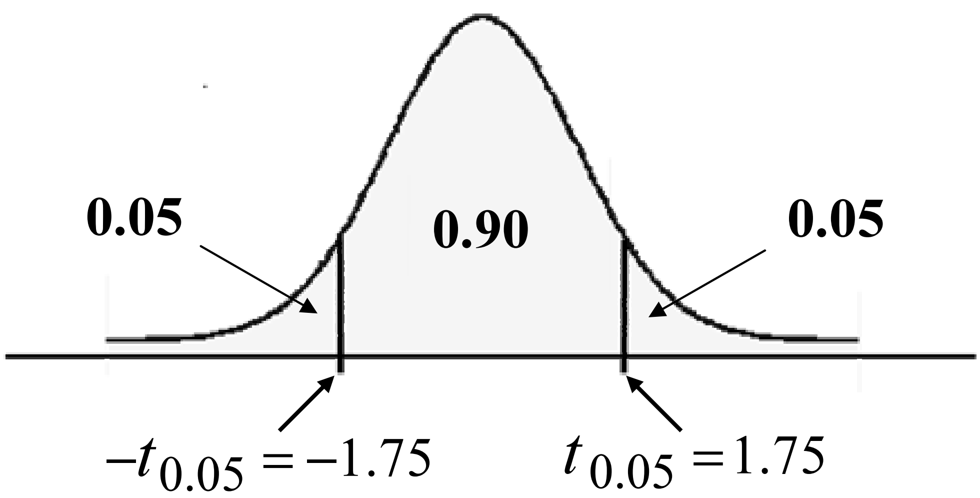
$$= [501.0319,\ 506.4681]$$

$$\left[\bar{X} - t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]$$

$$\bar{x} = 503.75\ ; \quad s = 6.2022\ ; \quad n = 16$$

$$1-\alpha = 0.90; \quad \alpha = 0.10; \quad \alpha/2 = 0.05 \ \Rightarrow$$

$$t_{\alpha/2;n-1} = t_{0.05;15} = 1.75$$



$$\left[503.75 - 1.75 \times \frac{6.2022}{4}, 503.75 + 1.75 \times \frac{6.2022}{4}\right] =$$

$$= [501.0319,\ 506.4681]$$

➢ 7.3.2.2. Intervalos de confianza para la varianza

❖ Media poblacional desconocida

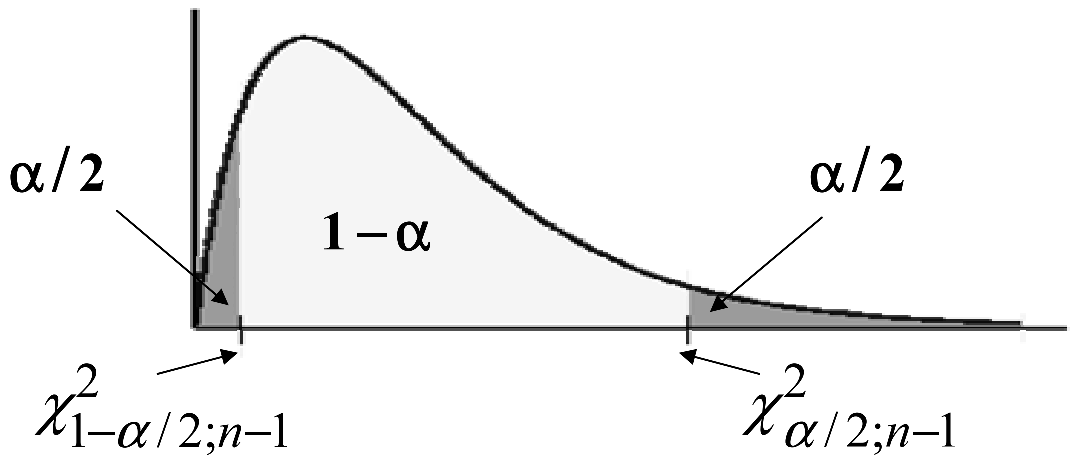
Estimador: Cuasivarianza $S^2 = \frac{1}{n-1}\sum_i \left(X_i - \bar{X}\right)^2$

⌘ Distribución muestral

$$\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \to \chi^2_{n-1}$$

$$P\left[\chi^2_{1-\alpha/2;n-1} \le \chi^2_{n-1} \le \chi^2_{\alpha/2;n-1}\right] = 1-\alpha$$

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



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

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

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

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

I.C. para σ^2, con media poblacional desconocida, al nivel de confianza $1-\alpha$

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}},\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}\right]$$

❑ **Ejemplo**

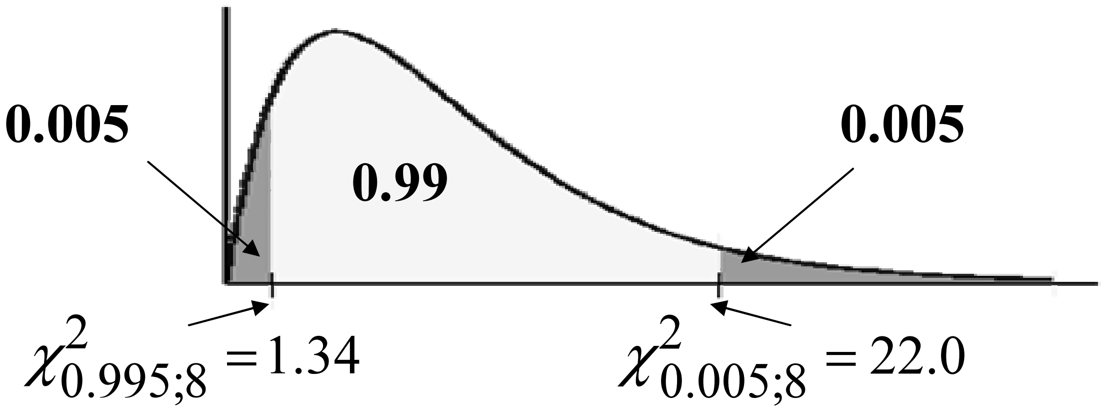
Se desea estimar la varianza del nivel de nistamina en un ungüento. Se conoce por larga experiencia que su distribución sigue una ley Normal. Se toma una muestra de 9 ungüentos, dando el nivel siguiente (en millones de unidades/gr): 1, 0.9, 1.5, 2.8, 3.1, 3.2, 2.5, 1.9, 2. Estimar la varianza mediante dos intervalos de confianza al nivel de confianza del 99% y del 95%.

Solución

Cuasivarianza: $$s^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2 = 0.74$$

Nivel de confianza: $1 - \alpha = 0.99$; $\alpha = 0.01$; $\alpha/2 = 0.005$

$$\chi^2_{\alpha/2;n-1} = \chi^2_{0.005;8} = 22.0; \chi^2_{1-\alpha/2;n-1} = \chi^2_{0.995;8} = 1.34$$



$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2;n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2;n-1}}\right] =$$

$$= [0.2691,\ 4.4179]$$

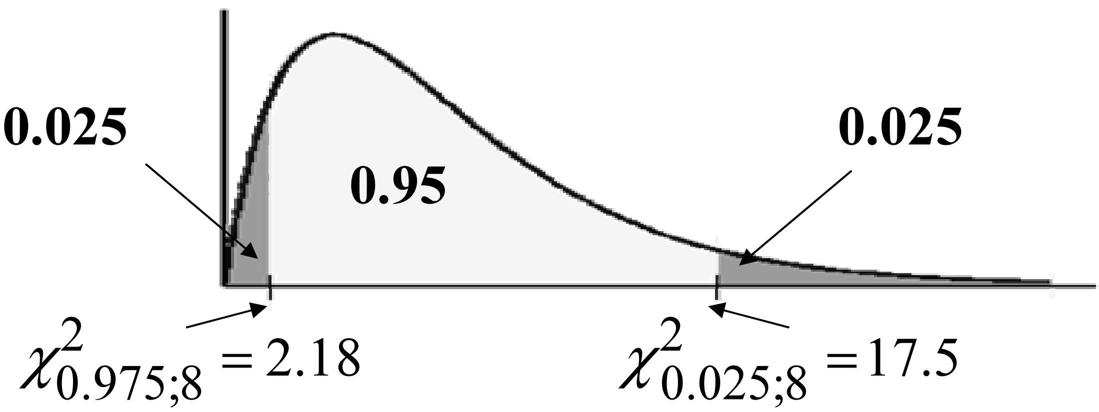
Nivel de confianza: $1-\alpha = 0.95$

$1-\alpha = 0.95; \quad \alpha = 0.05; \; \alpha/2 = 0.025$

Cuasi-varianza: $$s^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2 = 0.74$$

$$\chi^2_{\alpha/2\,;\,n-1} = \chi^2_{0.025\,;\,8} = 17.5$$

$$\chi^2_{1-\alpha/2\,;\,n-1} = \chi^2_{0.975\,;\,8} = 2.18$$



$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2;n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2;n-1}}\right] =$$

$$= [0.3383, \; 2.7156]$$

◈ 7.3.3. Intervalos de confianza para dos poblaciones Normales independientes

❖ *Se muestrean dos poblaciones normales para estimar los parámetros "comparativamente"*

Sean las variables aleatorias X e Y tales que

$$\left.\begin{array}{l} X \longrightarrow N(\mu_X; \sigma_X) \\ Y \longrightarrow N(\mu_Y; \sigma_Y) \end{array}\right\} \text{Independientes}$$

Consideramos:

$$\left.\begin{array}{c} \text{m.a.s. de tamaño } n_X \text{ de } X \\ X_1, X_2, \ldots, X_{n_X} \end{array}\right\} \overline{X},\ S_X^2$$

$$\left.\begin{array}{c} \text{m.a.s. de tamaño } n_Y \text{ de } Y \\ Y_1, Y_2, \ldots, Y_{n_Y} \end{array}\right\} \overline{Y},\ S_Y^2$$

➢ Se desean estimar comparativamente los parámetros de ambas poblaciones

➢ 7.3.3.1. Intervalos de confianza para la diferencia de medias

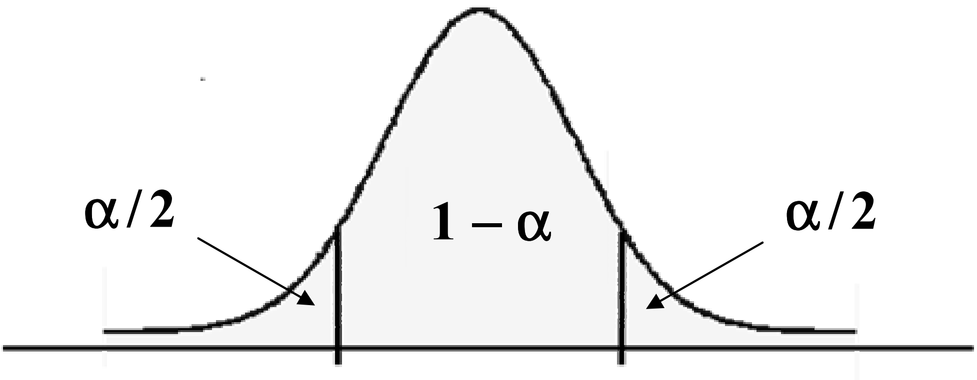
❖ Varianzas poblacionales conocidas

Estimador: $T = \overline{X} - \overline{Y}$

Distribución muestral:

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_Y\right)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \rightarrow N(0;1)$$

$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] = 1 - \alpha$$



$z_{1-\alpha/2} = -z_{\alpha/2}$ $\quad z_{\alpha/2}$

$$P\left[-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[-z_{\alpha/2} \leq \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sqrt{\dfrac{\sigma_X^2}{n_X}+\dfrac{\sigma_Y^2}{n_Y}}} \leq z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[\left(\bar{X}-\bar{Y}\right)-z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}} \leq \mu_X-\mu_Y \leq \right.$$

$$\left. \leq \left(\bar{X}-\bar{Y}\right)+z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\right] = 1-\alpha$$

I.C. para $\mu_X-\mu_Y$ con varianzas poblacionales conocidas, al nivel de confianza *1–α*

$$\left[\left(\bar{X}-\bar{Y}\right)\pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\right]$$

- **Ejemplo**

Se están utilizando normalmente en una granja avícola dos tipos de piensos compuestos A y B. Queriendo comparar la media de engorde con ambos piensos, para un nivel de confianza 0.9, se alimentan a 20 aves durante cierto tiempo con el pienso A obteniéndose una ganancia media de peso de 0.4 Kgr por ave. Simultáneamente a otras 19 aves se les alimenta con el pienso B y se obtiene un engorde medio de 0.5 Kgr.

Se conoce por experiencias previas que las variables objeto de estudio, engorde con cada uno de los piensos, son normales con varianzas de 0.05 para el pienso A y 0.1 para pienso B. Estimar la diferencia de engorde medio.

Solución

$$n_A = 20 \;; \quad \overline{x}_A = 0.4 \;; \quad \sigma_A^2 = 0.05$$

$$n_B = 19 \;; \quad \overline{x}_B = 0.5 \;; \quad \sigma_B^2 = 0.1$$

$$\left[\overline{x}_A - \overline{x}_B - z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}},\ \overline{x}_A - \overline{x}_B + z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \right]$$

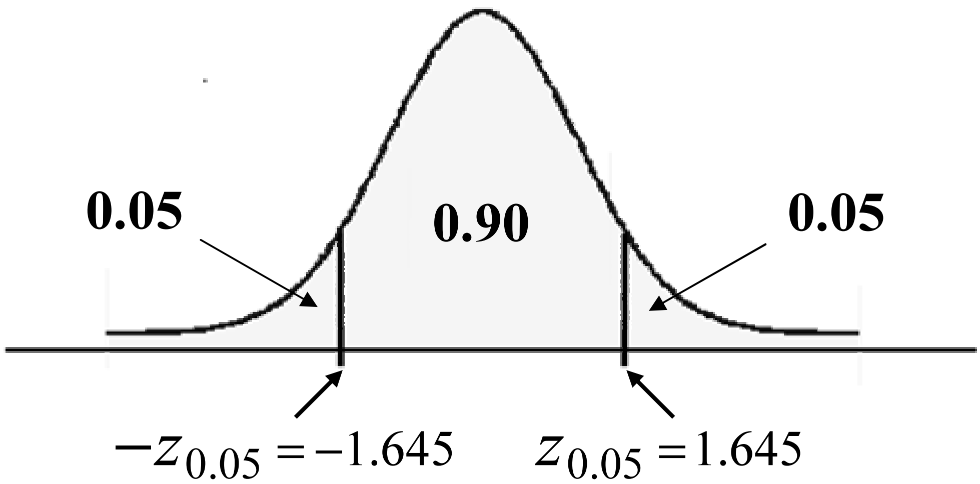
$$n_A = 20 \; ; \quad \bar{x}_A = 0.4 \; ; \quad \sigma_A^2 = 0.05$$

$$n_B = 19 \; ; \quad \bar{x}_B = 0.5 \; ; \quad \sigma_B^2 = 0.1$$

$$\left[\bar{x}_A - \bar{x}_B - z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}, \; \bar{x}_A - \bar{x}_B + z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \right]$$

$$1 - \alpha = 0.90; \quad \alpha = 0.10; \quad \alpha/2 = 0.05 \;\Rightarrow$$

$$z_{\alpha/2} = z_{0.05} = 1.645$$



$$\left[0.4 - 0.5 - 1.645\sqrt{\frac{0.05}{20} + \frac{0.1}{19}}, \; 0.4 - 0.5 + 1.645\sqrt{\frac{0.05}{20} + \frac{0.1}{19}} \right] =$$

$$= [-0.2449, \; 0.0449]$$

❖ Varianzas poblacionales desconocidas pero iguales $\sigma_X^2 = \sigma_Y^2$

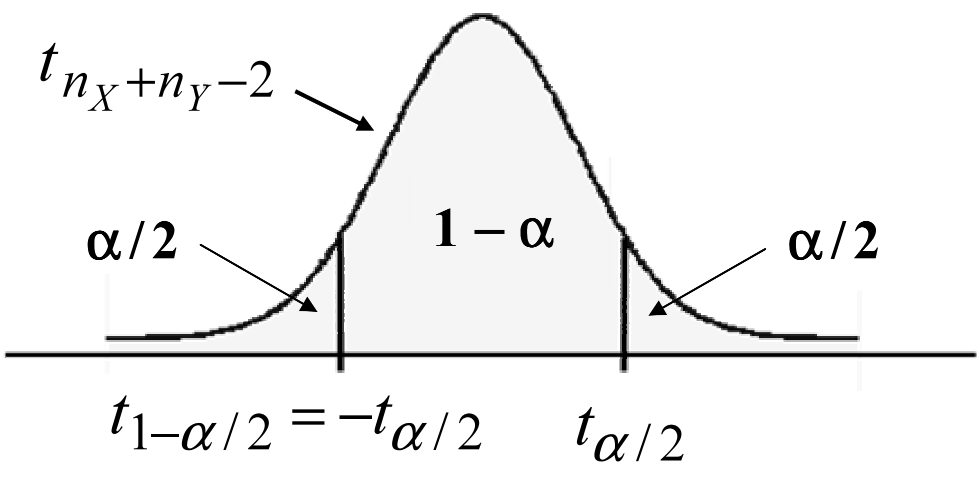
Estimador: $T = \bar{X} - \bar{Y}$

✦ Distribución muestral

$$T = \frac{\left(\bar{X} - \bar{Y}\right) - \left(\mu_x - \mu_y\right)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \rightarrow t_{n_x + n_y - 2}$$

$$S_p = \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}}$$

$$P\left[-t_{\alpha/2; n_X - n_Y - 2} \leq t_{n_X - n_Y - 2} \leq t_{\alpha/2; n_X - n_Y - 2}\right] = 1 - \alpha$$



$$P\left[-t_{\alpha/2;n_X-n_Y-2} \leq t_{n_X-n_Y-2} \leq t_{\alpha/2;n_X-n_Y-2}\right] = 1-\alpha$$

$$P\left[-t_{\alpha/2;\,n_X+n_Y-2} \leq \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{S_p\sqrt{\dfrac{1}{n_X}+\dfrac{1}{n_Y}}} \leq t_{\alpha/2;\,n_X+n_Y-2}\right] =$$

$$P\left[\bar{X}-\bar{Y}-t_{\alpha/2;\,n_X+n_Y-2}\times S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}} \leq \mu_X-\mu_Y \leq\right.$$

$$\left.\leq \bar{X}-\bar{Y}+t_{\alpha/2;\,n_X+n_Y-2}\times S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\right] = 1-\alpha$$

I.C. para $\mu_x - \mu_y$, al nivel de confianza $1-\alpha$, con varianzas poblacionales desconocidas pero iguales y muestra pequeñas

$$\left[\bar{X}-\bar{Y} \pm t_{\alpha/2;\,n_X+n_Y-2}\times S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\right]$$

❑ Ejemplo

Se están utilizando normalmente en una granja avícola dos tipos de piensos compuestos A y B. Queriendo comparar la media de engorde con ambos piensos, para un nivel de confianza 0.9, se alimentan a 22 aves durante cierto tiempo con el pienso A obteniéndose una ganancia media de peso de 0.4 Kgr por ave con una varianza de 0.03. Simultáneamente a otras 20 aves se les alimenta con el pienso B y se obtiene un engorde medio de 0.5 Kgr con una varianza de 0.09.

Se conoce por experiencias previas que las variables objeto de estudio, engorde con cada uno de los piensos, son normales con varianzas poblacionales iguales. Estimar la diferencia de engorde medio.

Solución

$$\left[\bar{x}_A - \bar{x}_B \pm t_{\alpha/2;n_A+n_B-2} \times s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\right]$$

$$n_A = 22\,; \quad \bar{x}_A = 0.4\,; \quad \hat{\sigma}_A^2 = 0.03$$

$$n_B = 20\,; \quad \bar{x}_B = 0.5\,; \quad \hat{\sigma}_B^2 = 0.09$$

$$n_A = 22\,; \quad \bar{x}_A = 0.4\,; \quad \hat{\sigma}_A^2 = 0.03$$

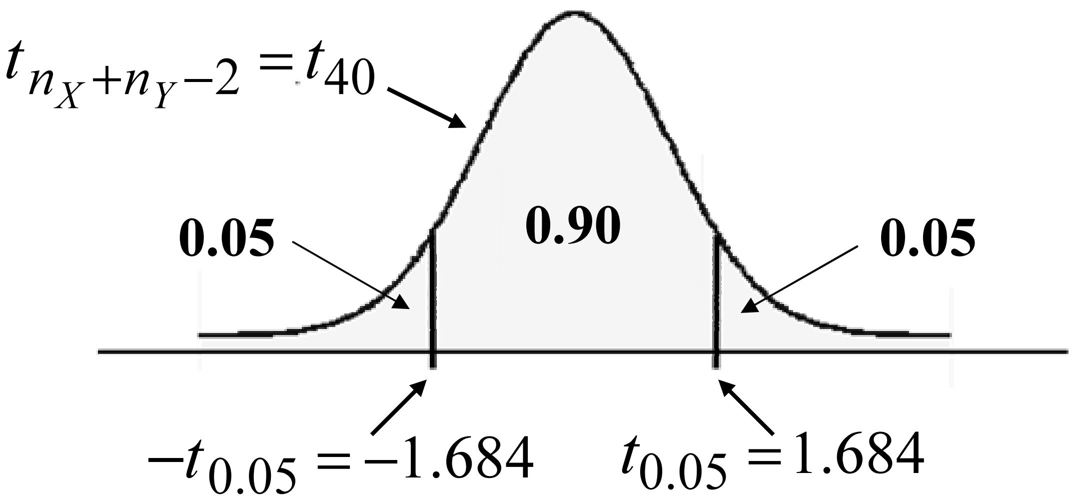
$$n_B = 20\,; \quad \bar{x}_B = 0.5\,; \quad \hat{\sigma}_B^2 = 0.09$$

$$s_A^2 = \frac{n_A}{n_A - 1}\,\hat{\sigma}_A^2 = 0.0316\,; \qquad s_B^2 = \frac{n_B}{n_B - 1}\,\hat{\sigma}_B^2 = 0.095$$

$$1-\alpha = 0.90; \quad \alpha = 0.10; \quad \alpha/2 = 0.05 \;\Rightarrow$$

$$t_{\alpha/2;n_A+n_B-2} = t_{0.05;40} = 1.684$$

$$s_p = \sqrt{\frac{(n_A-1)\,s_A^2 + (n_B-1)\,s_B^2}{n_A+n_B-2}} = 0.2484$$



$$\left[\bar{x}_A - \bar{x}_B \pm t_{\alpha/2;n_A+n_B-2} \times s_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}\right] =$$

$$= [-0.2292,\ 0.0292]$$

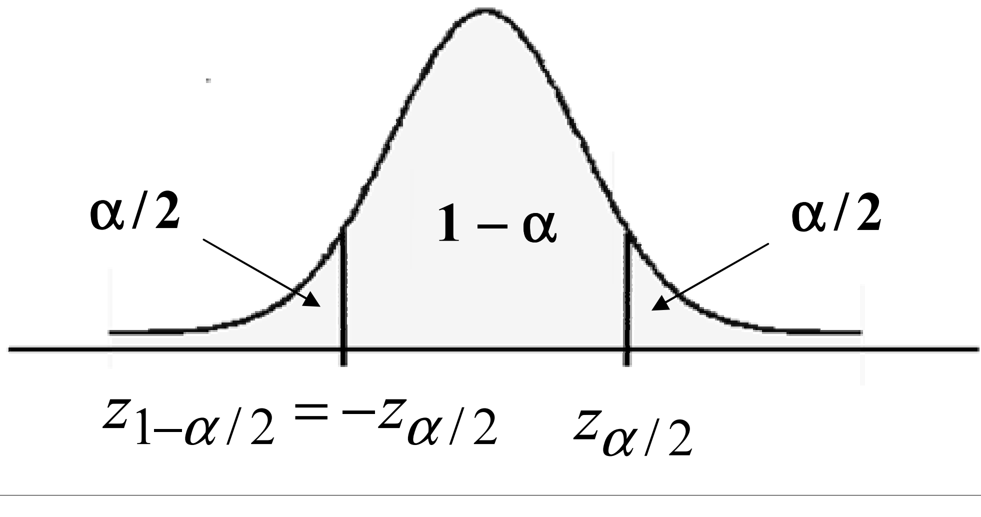
❖ Varianzas poblacionales desconocidas pero iguales, $\sigma_X^2 = \sigma_Y^2$ Tamaños muestrales grandes

Estimador: $T = \bar{X} - \bar{Y}$

✦ Distribución muestral:

$$Z = \frac{\left(\bar{X} - \bar{Y}\right) - \left(\mu_X - \mu_Y\right)}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}} \xrightarrow{n_X, n_Y \to \infty} N(0;\, 1)$$

$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] = 1 - \alpha$$



NOTA: La aproximación se considera correcta para

n_X y n_Y > 30

$$P\left[-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[-z_{\alpha/2} \leq \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}} \leq z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[\left(\bar{X}-\bar{Y}\right)-z_{\alpha/2}\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}} \leq \mu_X-\mu_Y \leq \right.$$

$$\left. \leq \left(\bar{X}-\bar{Y}\right)+z_{\alpha/2}\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}\right] = 1-\alpha$$

I.C. para $\mu_X - \mu_Y$, al nivel de confianza $1-\alpha$, con varianzas poblacionales desconocidas y muestras grandes

$$\left[\left(\bar{X}-\bar{Y}\right) \pm z_{\alpha/2}\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}\right]$$

- ❑ **Ejemplo**

Se están utilizando normalmente en una granja avícola dos tipos de piensos compuestos A y B. Queriendo comparar la media de engorde con ambos piensos, para un nivel de confianza 0.9, se alimentan a 100 aves durante cierto tiempo con el pienso A obteniéndose una ganancia media de peso de 0.5 Kgr por ave con una cuasivarianza de 0.08. Simultáneamente a otras 120 aves se les alimenta con el pienso B y se obtiene un engorde medio de 0.2 Kgr con una cuasivarianza de 0.09.
Estimar la diferencia de engorde medio.

Solución

$$\left[\left(\bar{x}_A - \bar{x}_B\right) \pm z_{\alpha/2}\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} \right]$$

$$n_A = 100\,; \quad \bar{x}_A = 0.5\,; \quad s_A^2 = 0.08$$

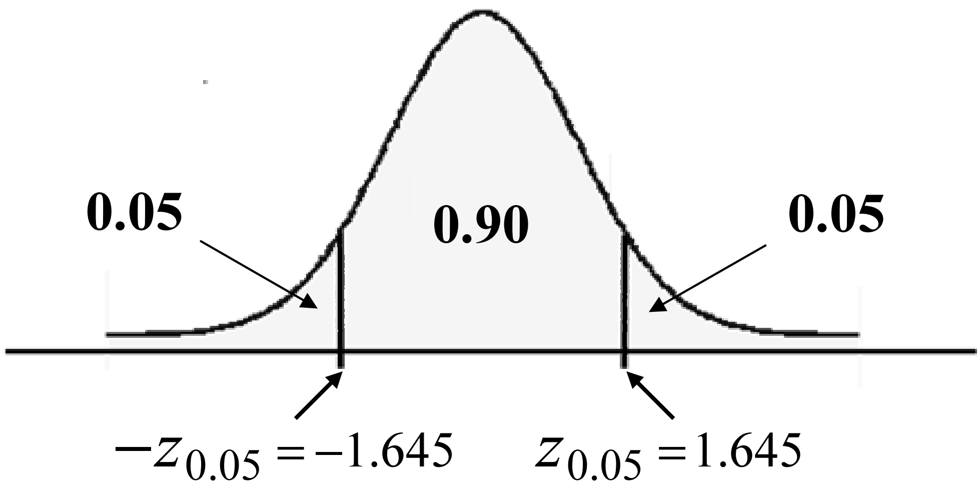
$$n_B = 120\,; \quad \bar{x}_B = 0.2\,; \quad s_B^2 = 0.09$$

$$n_A = 100\,;\quad \bar{x}_A = 0.5\,;\quad s_A^2 = 0.08$$

$$n_B = 120\,;\quad \bar{x}_B = 0.2\,;\quad s_B^2 = 0.09$$

$$1-\alpha = 0.90;\quad \alpha = 0.10;\quad \alpha/2 = 0.05 \;\Rightarrow$$

$$z_{\alpha/2} = z_{0.05} = 1.645$$



$$\left[\left(\bar{x}_A - \bar{x}_B\right) \pm z_{\alpha/2}\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}\right] =$$

$$= \left[0.5 - 0.2 - 1.645\sqrt{\frac{0.08}{100} + \frac{0.09}{120}},\ 0.5 - 0.2 + 1.645\sqrt{\frac{0.08}{100} + \frac{0.09}{120}}\right] =$$

$$= [0.2352,\ 0.3648]$$

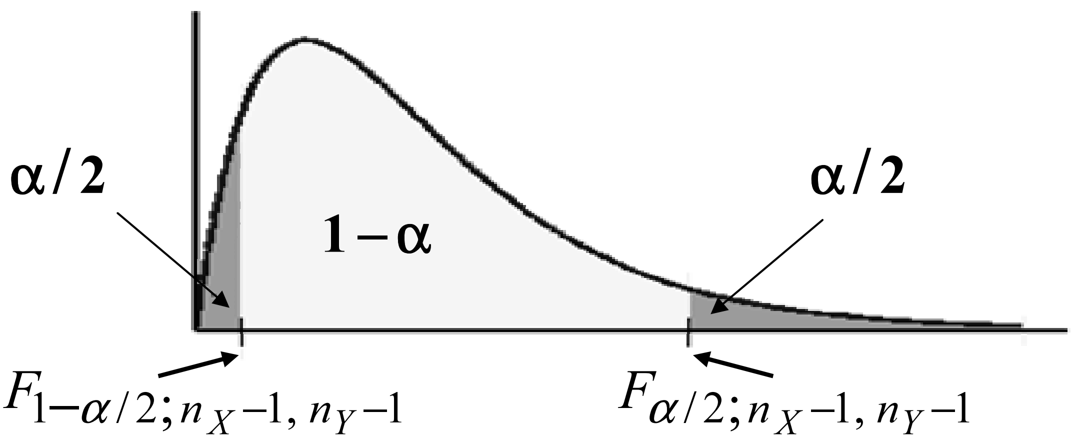
➢ 7.3.3.2. Intervalos de confianza para el cociente de varianzas: σ_X^2 / σ_Y^2

Estimador: $$T = \frac{S_X^2}{S_Y^2} = \frac{\frac{1}{n_X - 1}\sum_{i=1}^{n_X}\left(X_i - \overline{X}\right)^2}{\frac{1}{n_Y - 1}\sum_{i=1}^{n_Y}\left(Y_i - \overline{Y}\right)^2}$$

Distribución muestral

$$F = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} = \frac{S_X^2 \sigma_Y^2}{S_Y^2 \sigma_X^2} \rightarrow F_{n_X - 1,\, n_Y - 1}$$

$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1} \leq F_{n_X-1,n_Y-1} \leq F_{\alpha/2;n_X-1,n_Y-1}\right] = 1-\alpha$$



$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1} \le F_{n_X-1,n_Y-1} \le F_{\alpha/2;n_X-1,n_Y-1}\right] = 1-\alpha$$

$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1} \le \frac{\sigma_Y^2}{\sigma_X^2}\frac{S_X^2}{S_Y^2} \le F_{\alpha/2;n_X-1,n_Y-1}\right] = 1-\alpha$$

$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1}\frac{S_Y^2}{S_X^2} \le \frac{\sigma_Y^2}{\sigma_X^2} \le F_{\alpha/2;n_X-1,n_Y-1}\frac{S_Y^2}{S_X^2}\right] = 1-\alpha$$

$$P\left[\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2} \ge \frac{\sigma_X^2}{\sigma_Y^2} \ge \frac{S_X^2}{S_Y^2}\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}}\right] = 1-\alpha$$

$$P\left[\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2} \le \frac{\sigma_X^2}{\sigma_Y^2} \le \frac{S_X^2}{S_Y^2}\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\right] = 1-\alpha$$

$$P\left[\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2}\le\frac{\sigma_X^2}{\sigma_Y^2}\le\frac{S_X^2}{S_Y^2}\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\right]=1-\alpha$$

$$\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}=F_{\alpha/2;n_Y-1,n_X-1}$$

$$P\left[\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2}\le\frac{\sigma_X^2}{\sigma_Y^2}\le F_{\alpha/2;n_Y-1,n_X-1}\frac{S_X^2}{S_Y^2}\right]=1-\alpha$$

I.C. para σ_X^2 / σ_Y^2 con medias poblacionales desconocidas, al nivel de confianza $1-\alpha$

$$\left[\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2}\ ,\ F_{\alpha/2\,;\,n_Y-1\,,\,n_X-1}\frac{S_X^2}{S_Y^2}\right]$$

❑ **Ejemplo**

Una central lechera recibe diariamente leche de dos granjas A y B. Deseando estudiar la calidad de los productos se eligen dos muestras al azar de la leche suministrada por cada una de las granjas analizando el contenido en grasa. Para la granja A se han tomado 11 muestras obteniéndose una cuasivarianza de 0.034, mientras que para la granja B ha sido de 0.027 en un total de 16 muestras.

Es conocido por experiencias previas que los contenidos medios en grasa de la granjas son normales e independientes. Estimar el cociente de varianzas al nivel de confianza de 0.98.

Solución

$$\left[\frac{1}{F_{\alpha/2;n_A-1,n_B-1}}\frac{s_A^2}{s_B^2}\ ,\ F_{\alpha/2;n_B-1,n_A-1}\frac{s_A^2}{s_B^2}\right]$$

$$n_A = 11\ ;\ \ s_A^2 = 0.034\ ;\quad n_B = 16\ ;\ \ s_B^2 = 0.027$$

$$n_A = 11\ ;\ \ s_A^2 = 0.034\ ;\quad n_B = 16\ ;\ \ s_B^2 = 0.027$$

$$1-\alpha = 0.98;\quad \alpha = 0.02;\quad \alpha/2 = 0.01 \ \Rightarrow$$

$$F_{\alpha/2;\,n_A-1,\,n_B-1} = F_{0.01;10,15} = 3.80$$

$$F_{\alpha/2;\,n_B-1,\,n_A-1} = F_{0.01;15,10} = 4.56$$

$$\left[\frac{1}{F_{\alpha/2;n_A-1,n_B-1}}\frac{s_A^2}{s_B^2}\ ,\ F_{\alpha/2;n_B-1,n_A-1}\frac{s_A^2}{s_B^2}\right] =$$

$$= \left[\frac{1}{3.80}\times\frac{0.034}{0.027}\ ,\ 4.56\times\frac{0.034}{0.027}\right] = [0.3314,\ 5.7422]$$

◈ 7.3.4. Intervalo de confianza para una proporción

❖ *Se muestrea una población para estimar el parámetro proporción*

p : proporción de éxitos en la población

X : "número de éxitos en n realizaciones independientes"

n conocido

$$X \longrightarrow B\,(n\,;\,p)$$

Parámetro a estimar: p

Estimador puntual de p: $\hat{p} = X/n$

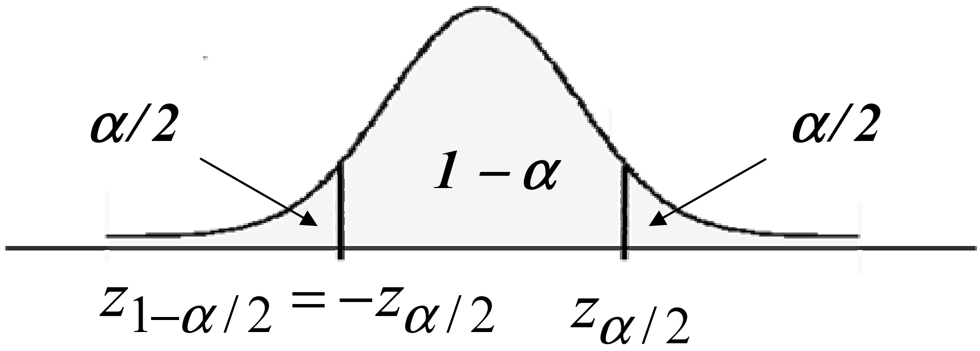
Distribución asintótica (n $\to \infty$):

$$X \xrightarrow{n \;\to\; \infty} N\left(np\,;\, n\hat{p}\left(1-\hat{p}\right)\right)$$

$$\hat{p} = \frac{X}{n} \xrightarrow{n \to \infty} N\left(p\,;\, \frac{\hat{p}\left(1-\hat{p}\right)}{n}\right)$$

$$Z = \frac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}\,(1-\hat{p}\,)}{n}}} \xrightarrow{n \;\to\; \infty} N\left(0;1\right)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\left(1 - \hat{p}\right)}{n}}} \xrightarrow{\ n \to \infty\ } N(0;1)$$



$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] = 1 - \alpha$$

$$P\left[-z_{\alpha/2} \le \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}} \le z_{\alpha/2}\right] = 1 - \alpha$$

$$P\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1 - \hat{p}\right)}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1 - \hat{p}\right)}{n}}\right] = 1 - \alpha$$

I.C. para el parámetro proporción, al nivel 1−α

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1 - \hat{p}\right)}{n}}\ ,\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1 - \hat{p}\right)}{n}}\right]$$

- **Ejemplo**

Se ignora la proporción de ranas tigre que se encuentran en una región de México. Para dar una estimación se ha tomado una muestra de 100 ranas observando que 15 de ellas son de este tipo. Hallar un intervalo de confianza al nivel de confianza del 0.95 para la proporción de ranas tigre.

Solución

$$\hat{p} = \frac{15}{100} = 0.15 \quad ; \quad z_{\alpha/2} = z_{0.025} = 1.96$$

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] =$$

$$= \left[0.15 - 1.96\sqrt{\frac{0.15 \times 0.85}{100}},\ 0.15 + 1.96\sqrt{\frac{0.15 \times 0.85}{100}}\right] =$$

$$= [0.081,\ 0.219]$$

◈ 7.3.5. Intervalo de confianza para la diferencia de proporciones

Se muestrean dos poblaciones independientes para estimar la diferencia de proporciones

$p_1 - p_2$: diferencia de proporciones de éxitos en la población

X : "número de éxitos en n realizaciones independientes"

Y : "número de éxitos en m realizaciones independientes"

n_X y n_Y conocidos

Parámetro a estimar : $p_1 - p_2$

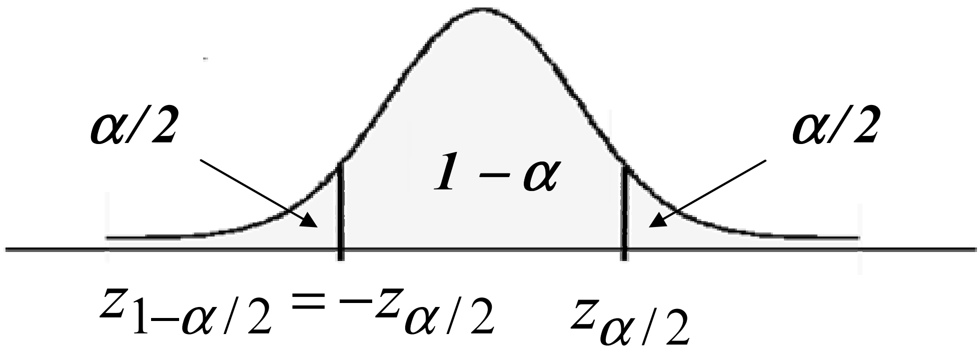
Estimador puntual de $p_1 - p_2$: $\widehat{p_1} - \widehat{p_2} = X/n_X - Y/n_Y$

Distribución asintótica $\left(n_X,\ n_Y \ \to\ \infty\right)$:

$$\widehat{p_1} - \widehat{p_2} = \frac{X}{n_X} - \frac{Y}{n_Y} \xrightarrow{n_X,\, n_Y \ \to\ \infty}$$

$$N\left(p_1 - p_2 ; \frac{\widehat{p_1}\left(1 - \widehat{p_1}\right)}{n_X} + \frac{\widehat{p_2}\left(1 - \widehat{p_2}\right)}{n_Y} \right)$$

$$Z = \frac{\left(\widehat{p_1} - \widehat{p_2}\right) - \left(p_1 - p_2\right)}{\sqrt{\frac{\widehat{p_1}\left(1 - \widehat{p_1}\right)}{n_X} + \frac{\widehat{p_2}\left(1 - \widehat{p_2}\right)}{n_Y}}} \xrightarrow{n_X, n_Y \to \infty} N(0;1)$$



$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] =$$

$$P\left(-z_{\alpha/2} \le \frac{\left(\widehat{p_1} - \widehat{p_2}\right) - \left(p_1 - p_2\right)}{\sqrt{\frac{\widehat{p_1}\left(1 - \widehat{p_1}\right)}{n_X} + \frac{\widehat{p_2}\left(1 - \widehat{p_2}\right)}{n_Y}}} \le z_{\alpha/2} \right) = 1 - \alpha$$

$$P\left[-z_{\alpha/2} \le \frac{\left(\widehat{p_1}-\widehat{p_2}\right)-\left(p_1-p_2\right)}{\sqrt{\dfrac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\dfrac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}}} \le \right]=1-\alpha$$

$$P\left[\left(\widehat{p_1}-\widehat{p_2}\right)-z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}} \le \left(p_1-p_2\right) \le \right.$$

$$\left. \le \left(\widehat{p_1}-\widehat{p_2}\right)+z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}}\right]=1-\alpha$$

I.C. para la diferencia de proporciones, al nivel de confianza $1-\alpha$

$$\left[\left(\widehat{p_1}-\widehat{p_2}\right)\pm z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}}\right]$$

❑ **Ejemplo**

Se desean comparar las proporciones de ranas pipiens que se encuentran en dos regiones independientes de México. Para dar una estimación se ha tomado una muestra de 80 ranas observando que 5 de ellas son de este tipo en la zona A, habiendo 8 de 100 en la zona B. Hallar un intervalo de confianza al nivel de confianza del 0.95 para la diferencia de proporciones de ranas pipiens.

Solución

$$\widehat{p_1} = \frac{5}{80} = 0.0625\ ; \qquad z_{\alpha/2} = z_{0.025} = 1.96$$

$$\widehat{p_2} = \frac{8}{100} = 0.08\ ; \qquad n_X = 80\ ; \qquad n_Y = 100$$

$$\left[\widehat{p_1} - \widehat{p_2} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X} + \frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}}\right] =$$

$$\left[0.0625 - 0.08 \pm 1.96\sqrt{\frac{0.0625(1-0.0625)}{80} + \frac{0.08(1-0.08)}{100}}\right] =$$

$$= [-0.0926,\ 0.0576]$$