TEMA 4: ESTIMACIÓN DE PARÁMETROS MEDIANTE INTERVALOS DE CONFIANZA

(Bibliografía básica: *"Técnicas Cuantitativas para la Inferencia"* y *"Ejercicios Resueltos de Técnicas Cuantitativas para la Inferencia"*)

1. CONCEPTO DE INTERVALO DE CONFIANZA PARA UN PARÁMETRO POBLACIONAL

X θ desconocido
↓
n → (X_1, X_2, ..., X_n)
↓
$\underset{estimador}{\hat{\theta}}$ = t(X_1, X_2, ..., X_n)

Hasta ahora cuando teníamos un parámetro poblacional θ desconocido lo estimábamos a través de un estimador puntual, $\hat{\theta}$.

Por ejemplo, la media poblacional la estimábamos a través de la media muestral, $\overline{X}$.

X θ desconocido
↓
n → (X_1, X_2, ..., X_n)
↙ ↘
$T_1 = t_1(X_1, X_2, ..., X_n)$ (extremo inferior) $T_2 = t_2(X_1, X_2, ..., X_n)$ (extremo superior)

En este tema lo que se pretende es construir un intervalo de confianza, para lo cual se fija una probabilidad β de que el parámetro pertenezca al intervalo y, basándonos en una m. a. s., obtenemos los extremos del intervalo T_1 y T_2:

Los extremos del intervalo $T_1 = t_1(X_1, X_2, ..., X_n)$ y $T_2 = t_2(X_1, X_2, ..., X_n)$ son funciones de las observaciones muestrales, las cuales son variables aleatorias, por lo que los extremos también tienen el carácter de variables aleatorias.

Por ello, la posibilidad de que el extremo inferior sea menor que θ y el extremo superior sea mayor que θ, tiene el carácter de probabilidad, que se ha denominado probabilidad β.

$$P[t_1(X_1, X_2, ..., X_n) < \theta < t_2(X_1, X_2, ..., X_n)] = \beta = 1 - \alpha$$

o $$P[T_1 < \theta < T_2] = \beta = 1 - \alpha$$

A la probabilidad de que el parámetro esté incluido dentro del intervalo (pertenezca al intervalo) se le llama β: probabilidad fiduciaria o nivel de confianza.

A la probabilidad de que el parámetro esté fuera del intervalo (no pertenezca al intervalo que se ha construido) se le llama α: riesgo de error o nivel de significación.

Cualquier valor del intervalo (T_1, T_2) puede tomarse como estimación puntual del parámetro poblacional θ desconocido.

Vamos a construir los intervalos de confianza para ciertos parámetros poblacionales (media, varianza, diferencia de medias y cociente de varianzas), suponiendo siempre poblaciones normales, excepto en el epígrafe 6, en el que se trabajará con poblaciones binomiales (entonces construiremos intervalos de confianza para proporciones poblacionales).

En todos los casos, para una probabilidad β prefijada, el intervalo de confianza que vamos a construir tendrá una amplitud mínima.

2. INTERVALO DE CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL

Vamos a distinguir los siguientes casos:

2.1. I. C. para la media de una población normal con varianza conocida.

2.2. I. C. para la media de una población normal con varianza desconocida:

a) Muestra pequeña $(n \leq 30)$.

b) Muestra grande $(n > 30)$.

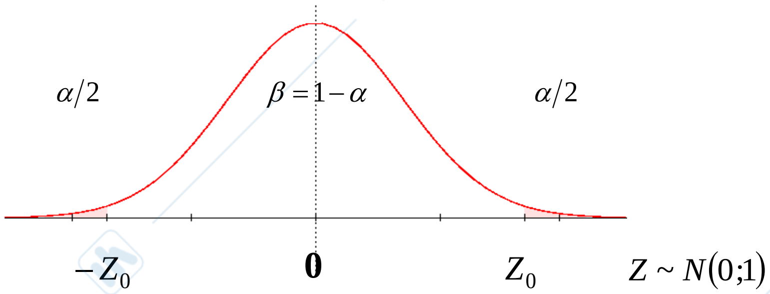
2.1. I. C. PARA LA MEDIA DE UNA POBLACIÓN NORMAL CON VARIANZA CONOCIDA

$$X \sim N(\mu;\sigma^2) \quad (\text{con } \sigma^2 \text{ conocida})$$
$$\downarrow \qquad \text{¿I. C. } (\mu)?$$
$$n \to (X_1, X_2, \ldots, X_n)$$
$$\downarrow$$
$$\bar{X} \sim N\left(\mu;\frac{\sigma^2}{n}\right) \to \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = Z \sim N(0;1)$$

(Tema 2. 1)

Una vez fijada la probabilidad $\beta = 1-\alpha$, establecemos que:

$$P(-Z_0 < Z < Z_0) = \beta = 1-\alpha$$



(Z_0 es el cuantil $1-\frac{\alpha}{2}$ de una $Z \sim N(0;1)$, por lo que $P(Z \le Z_0) = 1-\frac{\alpha}{2}$)

($-Z_0$ es el cuantil $\frac{\alpha}{2}$ de una $Z \sim N(0;1)$, y es el opuesto de Z_0)

A continuación se sustituye:

$$P\left(-Z_0 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < Z_0\right) = \beta = 1-\alpha$$

Se quiere construir el I. C. para el parámetro media poblacional, μ, por lo que hay que realizar operaciones para dejar sólo el parámetro en la parte central de la desigualdad.

Para ello, multiplicamos en los tres miembros de la desigualdad por $\frac{\sigma}{\sqrt{n}}$:

$$P\left(-Z_0\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < Z_0\frac{\sigma}{\sqrt{n}}\right) = \beta = 1-\alpha$$

Restamos en los tres miembros $\overline{X}$:

$$P\left(-\overline{X}-Z_0\frac{\sigma}{\sqrt{n}}<-\mu<-\overline{X}+Z_0\frac{\sigma}{\sqrt{n}}\right)=\beta=1-\alpha$$

Se multiplican los tres miembros por (-1), por lo que los signos de desigualdad cambian de sentido:

$$P\left(\overline{X}+Z_0\frac{\sigma}{\sqrt{n}}>\mu>\overline{X}-Z_0\frac{\sigma}{\sqrt{n}}\right)=\beta=1-\alpha$$

Se reordena el intervalo, de tal forma que el extremo inferior quede en la parte izquierda de la expresión y el extremo superior quede a la derecha:

$$\boxed{P\left(\overline{X}-Z_0\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+Z_0\frac{\sigma}{\sqrt{n}}\right)=\beta=1-\alpha}$$

La expresión del recuadro tiene una doble interpretación, pudiendo ambas coexistir en la práctica:

1ª) $\mu\in\left(\overline{X}-Z_0\frac{\sigma}{\sqrt{n}}\ ,\ \overline{X}+Z_0\frac{\sigma}{\sqrt{n}}\right)$ con una probabilidad $\beta=1-\alpha$.

Ésta va a ser la interpretación que usualmente se utilizará a lo largo del curso.

2ª) Se toman muestras sucesivas de tamaño n, y para cada una de las muestras se construye su correspondiente I. C. (los I. C. así obtenidos, previsiblemente, pueden ser diferentes, al ir cambiando las observaciones muestrales de una muestra a otra). Entonces, en el $\beta\cdot 100\,\%$ de los intervalos obtenidos, el verdadero valor de μ debe caer dentro de dichos intervalos obtenidos.

Puede observarse en la expresión del recuadro, que reproducimos de nuevo,

$$\boxed{P\left(\overline{X}-Z_0\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+Z_0\frac{\sigma}{\sqrt{n}}\right)=\beta=1-\alpha}$$

que se ha construido un intervalo simétrico cuyo centro es la media muestral, $\overline{X}$, y cuyos extremos se obtienen sumando y restando una misma cantidad, $Z_0\frac{\sigma}{\sqrt{n}}$, que es, por tanto, el radio del intervalo, que denominaremos r (observe que es un intervalo simétrico respecto de $\overline{X}$). Así, la expresión del radio es:

$$r = Z_0 \frac{\sigma}{\sqrt{n}}$$

$$\mu \in \left(\underbrace{\overline{X} - Z_0 \frac{\sigma}{\sqrt{n}} \;\; ---------- \;\; \overline{X}}_{Z_0 \frac{\sigma}{\sqrt{n}}} \underbrace{---------- \;\; \overline{X} + Z_0 \frac{\sigma}{\sqrt{n}}}_{Z_0 \frac{\sigma}{\sqrt{n}}}\right) \text{ con una probabilidad } \beta = 1 - \alpha$$

$$r = Z_0 \frac{\sigma}{\sqrt{n}}$$

Atendiendo a la primera interpretación, dicho intervalo contiene en su interior a la media poblacional, μ, con una probabilidad $\beta = 1 - \alpha$, que es el nivel de confianza del intervalo.

Por otro lado, el radio del intervalo, $r = Z_0 \frac{\sigma}{\sqrt{n}}$, también se puede interpretar como la cota superior (tope o límite máximo) para el error de estimación, siendo éste el error cometido al estimar μ mediante el estimador $\overline{X}$:

Error de estimación: $\left|\overline{X} - \mu\right|$

La interpretación del radio, r, como la cota superior o límite máximo para el error de estimación se deduce de la expresión del recuadro:

$$\boxed{P\left(\overline{X} - Z_0 \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_0 \frac{\sigma}{\sqrt{n}}\right) = \beta = 1 - \alpha}$$

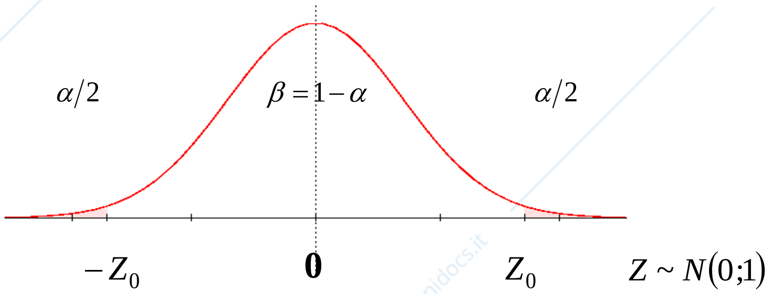
yendo hacia atrás:

$$P\left(\underbrace{-Z_0 \frac{\sigma}{\sqrt{n}}}_{=-r} < \overline{X} - \mu < \underbrace{Z_0 \frac{\sigma}{\sqrt{n}}}_{=r}\right) = \beta = 1 - \alpha$$

$$P\left(\underbrace{\left|\overline{X} - \mu\right|}_{\substack{\text{error de} \\ \text{estimación}}} < \underbrace{Z_0 \frac{\sigma}{\sqrt{n}}}_{=r}\right) = \beta = 1 - \alpha$$

el error de estimación es inferior al radio, r (que actúa, por tanto, como cota superior o límite máximo para el error de estimación), con una probabilidad $\beta = 1 - \alpha$.

Por otro lado, en la práctica se comprobará que cuanto mayor sea el nivel de confianza exigido, β, más alto es el valor del cuantil, Z_0, y mayor se hace el radio del intervalo, r (permaneciendo n constante):



$$\uparrow \beta \qquad \uparrow Z_0 \qquad \uparrow \left(r = Z_0 \frac{\sigma}{\sqrt{n}} \right) \qquad (\text{con } n \text{ constante})$$

Por el contrario, a medida que aumenta el tamaño de la muestra, n, el radio, r (cota superior o límite máximo para el error de estimación) disminuye, con lo que la estimación por medio del intervalo aumenta su precisión (permaneciendo β constante):

$$\uparrow n \qquad \downarrow \left(r = Z_0 \frac{\sigma}{\sqrt{n}} \right) \qquad (\text{con } \beta \text{ constante})$$

Si se pretende determinar el tamaño de la muestra que es necesario tomar para estimar μ, con un límite o cota para el error de estimación de cuantía r, prefijada, y con un nivel de confianza β, también dado, se despeja n de la expresión del radio, r:

$$\text{Dados } r \text{ y } \beta: \; r = Z_0 \frac{\sigma}{\sqrt{n}} \; ; \; \sqrt{n} = Z_0 \frac{\sigma}{r} \; ; \; \left(\sqrt{n}\right)^2 = \left(Z_0 \frac{\sigma}{r} \right)^2 \; ; \boxed{n = \left(\frac{Z_0 \, \sigma}{r} \right)^2}$$

Este es el tamaño de la muestra que ha de tomarse para estimar μ con un error de estimación que sea inferior a la cota o límite r, y con una probabilidad β.

2.2. I. C. PARA LA MEDIA DE UNA POBLACIÓN NORMAL CON VARIANZA DESCONOCIDA

a) Muestra pequeña (n ≤ 30)

$X \sim N(\mu;\sigma^2)$ $\quad$ (con σ^2 desconocida)

$\downarrow$ $\qquad$ ¿I. C. (μ)?

$n \rightarrow (X_1, X_2, \ldots, X_n)$ $\quad$ (con n ≤ 30)

$\swarrow \quad \downarrow \quad \searrow$

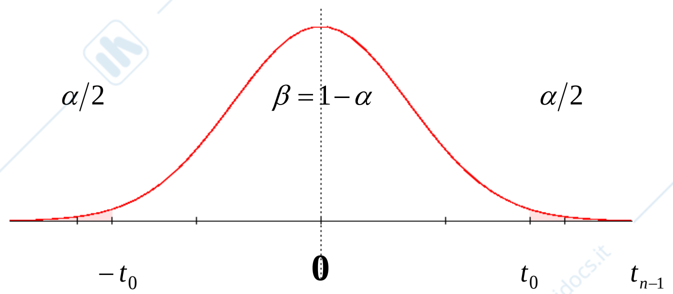
$\overline{X} \quad S_n^2 \quad S_{n-1}^2$

$$\left.\begin{array}{r}\sigma^2 \text{ desconocida} \\ n \leq 30\end{array}\right\}: \frac{\overline{X}-\mu}{S_n/\sqrt{n-1}} = \frac{\overline{X}-\mu}{S_{n-1}/\sqrt{n}} = t_{n-1}$$

(T. 2. 3. a)

Una vez fijado el nivel de confianza $\beta = 1-\alpha$ establecemos que:

$$P(-t_0 < t_{n-1} < t_0) = \beta = 1-\alpha$$



(t_0 es el cuantil $1-\dfrac{\alpha}{2}$ de una t_{n-1}, por lo que $P(t_{n-1} \leq t_0) = 1-\dfrac{\alpha}{2}$)

S_{n-1}) $\qquad P\left(-t_0 < \dfrac{\overline{X}-\mu}{S_{n-1}/\sqrt{n}} < t_0\right) = \beta = 1-\alpha$ $\qquad$ ó

S_n) $\qquad P\left(-t_0 < \dfrac{\overline{X}-\mu}{S_n/\sqrt{n-1}} < t_0\right) = \beta = 1-\alpha$

Se quiere construir un I. C. para μ, por lo que, haciendo operaciones similares a las del apartado 2.1., se obtiene:

S_{n-1}) $\qquad \boxed{P\left(\overline{X} - t_0 \dfrac{S_{n-1}}{\sqrt{n}} < \mu < \overline{X} + t_0 \dfrac{S_{n-1}}{\sqrt{n}}\right) = \beta = 1-\alpha}$ (con $r = t_0 \dfrac{S_{n-1}}{\sqrt{n}}$) ó

S_n) $$\boxed{P\left(\bar{X}-t_0\frac{S_n}{\sqrt{n-1}}<\mu<\bar{X}+t_0\frac{S_n}{\sqrt{n-1}}\right)=\beta=1-\alpha}$$ (con $r=t_0\frac{S_n}{\sqrt{n-1}}$)

- El tamaño muestral necesario para estimar μ, con un límite o cota para el error de estimación de cuantía *r* (con un nivel de confianza $\beta=1-\alpha$, dado), se obtiene despejando *n* de cualquiera de las dos expresiones anteriores de *r*:

$$r=t_0\frac{S_{n-1}}{\sqrt{n}}\ ;\quad \boxed{n=\left(\frac{\overset{Z_0}{t_0}\,S_{n-1}}{r}\right)^2} \qquad \text{ó} \qquad r=t_0\frac{S_n}{\sqrt{n-1}}\ ;\quad \boxed{n=\left(\frac{\overset{Z_0}{t_0}\,S_n}{r}\right)^2+1}$$

- Se utiliza el cuantil Z_0, en vez de t_0.
- Se utilizan S_{n-1} o S_n, obtenidas mediante un muestreo realizado con anterioridad al actual. Por tanto, es necesario un muestreo en 2 etapas (muestreo bietápico). Más detalles sobre la obtención de *n* en página 80 de *"Técnicas Cuantitativas para la Inferencia"*.

b) Muestra grande (n > 30)

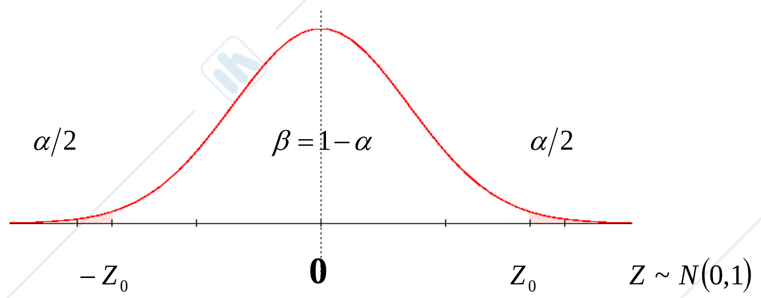
Recuerde que cuando n > 30: $t_n \xrightarrow[(n>30)]{} Z \sim N(0,1)$

Se vio también con anterioridad: $$t_{n-1}=\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}}=\frac{\bar{X}-\mu}{S_{n-1}/\sqrt{n}}\xrightarrow[(n>30)]{} Z\sim N(0,1)$$

(T. 2. 3. b)

Una vez fijado el nivel de confianza $\beta=1-\alpha$ establecemos que:

$$P(-Z_0<Z<Z_0)=\beta=1-\alpha$$



(Z_0 es el cuantil $1-\frac{\alpha}{2}$ de una $Z\sim N(0,1)$, por lo que $P(Z\le Z_0)=1-\frac{\alpha}{2}$)

S_{n-1}) $\quad P\left(-Z_0 < \dfrac{\overline{X}-\mu}{S_{n-1}/\sqrt{n}} < Z_0\right) = \beta = 1-\alpha$ ó

S_n) $\quad P\left(-Z_0 < \dfrac{\overline{X}-\mu}{S_n/\sqrt{n-1}} < Z_0\right) = \beta = 1-\alpha$

Se quiere construir un I. C. para μ, por lo que, haciendo operaciones similares a las del apartado 2.1., se llega a:

S_{n-1}) $\boxed{P\left(\overline{X}-Z_0\dfrac{S_{n-1}}{\sqrt{n}} < \mu < \overline{X}+Z_0\dfrac{S_{n-1}}{\sqrt{n}}\right) = \beta = 1-\alpha}$ (con $r = Z_0\dfrac{S_{n-1}}{\sqrt{n}}$) ó

S_n) $\boxed{P\left(\overline{X}-Z_0\dfrac{S_n}{\sqrt{n-1}} < \mu < \overline{X}+Z_0\dfrac{S_n}{\sqrt{n-1}}\right) = \beta = 1-\alpha}$ (con $r = Z_0\dfrac{S_n}{\sqrt{n-1}}$)

- El tamaño muestral necesario para estimar μ, con un límite o cota para el error de estimación de cuantía *r* (con un nivel de confianza $\beta = 1-\alpha$, dado), se obtiene despejando *n* de cualquiera de las dos expresiones anteriores de *r*:

$r = Z_0\dfrac{S_{n-1}}{\sqrt{n}}$; $\boxed{n = \left(\dfrac{Z_0\, S_{n-1}}{r}\right)^2}$ ó $r = Z_0\dfrac{S_n}{\sqrt{n-1}}$; $\boxed{n = \left(\dfrac{Z_0\, S_n}{r}\right)^2 + 1}$

- Para poder calcular *n* ha de utilizarse S_{n-1} o S_n, obtenidas mediante un muestreo realizado con anterioridad al actual. Por tanto, es necesario un muestreo bietápico.

3. I. C. PARA LA VARIANZA DE UNA POBLACIÓN NORMAL

3.1. Intervalo de confianza bilateral (véanse páginas 97-99 de *"Técnicas Cuantitativas para la Inferencia"*).

Caso A: μ y σ^2 desconocidas

Caso B: μ conocida σ^2 desconocida

<u>Caso A</u>: μ y σ^2 desconocidas

$X \sim N(\mu;\sigma^2)$ (con μ y σ^2 desconocidas)

¿I. C. (σ^2)?

$n \rightarrow (X_1, X_2, \ldots, X_n)$

$$\bar{X} \qquad S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2 \qquad S_{n-1}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

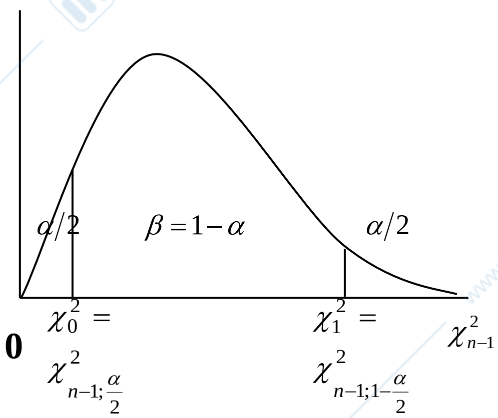
Recuerde que:

$$\frac{n S_n^2}{\sigma^2} = \frac{(n-1)S_{n-1}^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \chi_{n-1}^2$$

(T. 2. 2)

Una vez fijado el nivel de confianza $\beta = 1-\alpha$ establecemos que:

$$P(\chi_0^2 < \chi_{n-1}^2 < \chi_1^2) = \beta = 1-\alpha$$



(χ_0^2 es el cuantil $\alpha/2$ de una χ_{n-1}^2, por lo que $P(\chi_{n-1}^2 \le \chi_0^2) = \frac{\alpha}{2}$)

(χ_1^2 es el cuantil $1-\frac{\alpha}{2}$ de una χ_{n-1}^2, por lo que $P(\chi_{n-1}^2 \le \chi_1^2) = 1-\frac{\alpha}{2}$)

S_n^2) $\quad P\left(\chi_0^2 < \frac{n S_n^2}{\sigma^2} < \chi_1^2\right) = \beta = 1-\alpha$ ó

S_{n-1}^2) $\quad P\left(\chi_0^2 < \frac{(n-1)S_{n-1}^2}{\sigma^2} < \chi_1^2\right) = \beta = 1-\alpha$

Se quiere construir un I. C. para σ^2, por lo que se hacen operaciones en el intervalo del estadístico $\frac{nS_n^2}{\sigma^2}$ (o bien, en el del estadístico $\frac{(n-1)S_{n-1}^2}{\sigma^2}$):

se divide en los tres miembros de las desigualdades por nS_n^2:

$$P\left(\frac{\chi_0^2}{nS_n^2} < \frac{1}{\sigma^2} < \frac{\chi_1^2}{nS_n^2}\right) = \beta = 1-\alpha$$

se calcula el inverso de los tres miembros (las desigualdades cambian de sentido):

$$P\left(\frac{nS_n^2}{\chi_0^2} > \sigma^2 > \frac{nS_n^2}{\chi_1^2}\right) = \beta = 1-\alpha$$

se reordena:

S_n^2) $\boxed{P\left(\frac{nS_n^2}{\chi_1^2} < \sigma^2 < \frac{nS_n^2}{\chi_0^2}\right) = \beta = 1-\alpha}$ ó

S_{n-1}^2) $\boxed{P\left(\frac{(n-1)S_{n-1}^2}{\chi_1^2} < \sigma^2 < \frac{(n-1)S_{n-1}^2}{\chi_0^2}\right) = \beta = 1-\alpha}$

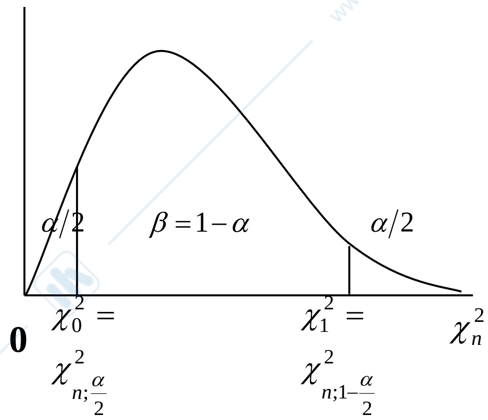
<u>Caso B</u>: μ conocida y σ^2 desconocida

Se utiliza el estadístico

$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$, que sigue una distribución χ_n^2, como se vio en T. 2. 2 (la demostración también se encuentra en la página 66 de *"Técnicas Cuantitativas para la Inferencia"* y en el Ejercicio 2 de la "Relación de Ejercicios Complementarios").

Se parte de que:

$$P\left(\chi_0^2 < \chi_n^2 < \chi_1^2\right) = \beta = 1-\alpha$$



$$P\left(\chi_0^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} < \chi_1^2\right) = \beta = 1-\alpha$$

y haciendo operaciones análogas a las del caso A (véase página 152 de "*Ejercicios resueltos de técnicas cuantitativas para la inferencia*"), se llega a:

$$\boxed{P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi_1^2} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi_0^2}\right) = \beta = 1-\alpha}$$

χ_1^2 y χ_0^2 son cuantiles de la distribución χ_n^2 (véase la gráfica de arriba).

3.2. Intervalo de confianza unilateral (véanse páginas 92-96 de *"Técnicas Cuantitativas para la Inferencia"*).

En ocasiones interesa conocer una cota superior de la varianza, es decir, qué valor de la varianza es difícil de sobrepasar, para lo cual, se construye un I. C. unilateral (en lugar de los I. C. bilaterales que se han construido hasta ahora).

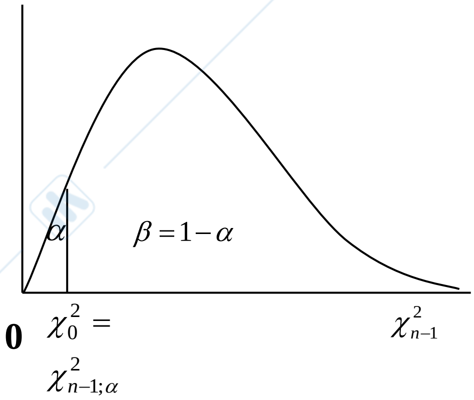
Caso A: μ y σ^2 desconocidas

Se utiliza el estadístico:

$$\frac{nS_n^2}{\sigma^2} = \frac{(n-1)S_{n-1}^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} = \chi_{n-1}^2$$

Una vez fijado el nivel de confianza $\beta = 1-\alpha$ establecemos que:

$$P\left(\chi^2_{n-1} > \chi^2_0\right) = \beta = 1-\alpha$$



(χ^2_0 es el cuantil α de una χ^2_{n-1}, por lo que $P\left(\chi^2_{n-1} \leq \chi^2_0\right) = \alpha$)

$$P\left(\frac{nS_n^2}{\sigma^2} > \chi^2_0\right) = \beta = 1-\alpha$$

Se hacen operaciones análogas a las del apartado 3. 1. A:

$$P\left(\frac{1}{\sigma^2} > \frac{\chi^2_0}{nS_n^2}\right) = \beta = 1-\alpha$$

S_n^2) $\boxed{P\left(\sigma^2 < \frac{nS_n^2}{\chi^2_0}\right) = \beta = 1-\alpha}$ ó

S_{n-1}^2) $\boxed{P\left(\sigma^2 < \frac{(n-1)S_{n-1}^2}{\chi^2_0}\right) = \beta = 1-\alpha}$ o $\boxed{P\left(\sigma^2 < \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\chi^2_0}\right) = \beta = 1-\alpha}$

<u>Caso B</u>: μ conocida y σ^2 desconocida

Se utiliza el estadístico

$\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\sigma^2}$, que sigue una distribución χ^2_n, y con un planteamiento análogo al del Caso A anterior, se obtiene:

$$\boxed{P\left(\sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi_0^2}\right) = \beta = 1-\alpha}$$

χ_0^2 es el cuantil α de la distribución χ_n^2, por lo que $P\left(\chi_n^2 \leq \chi_0^2\right) = \alpha$

4. I. C. PARA LA DIFERENCIA DE MEDIAS POBLACIONALES

Se pueden distinguir los cuatro casos que se derivan de lo explicado en el Tema 2.4.

4.1. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales conocidas.

4.2. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales desconocidas y muestras grandes (n+m>30).

4.3. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales desconocidas, pero iguales, y muestras pequeñas (n+$m\leq$30).

4.4. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales desconocidas y desiguales, y muestras pequeñas (n+$m\leq$30).

4.1. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales conocidas.

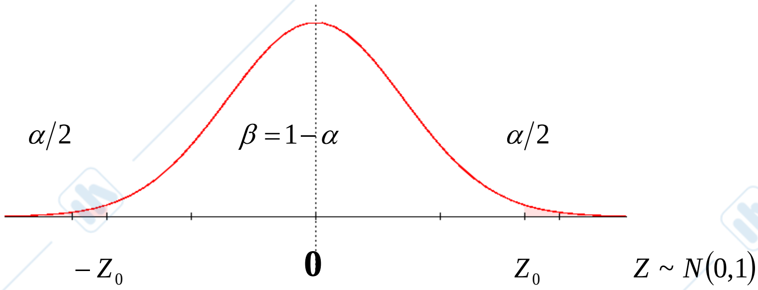
$X \sim N\left(\mu_1;\sigma_1^2\right)$	σ_1^2 , σ_2^2 conocidas	$Y \sim N\left(\mu_2;\sigma_2^2\right)$
$\downarrow$		$\downarrow$
$n \rightarrow \left(X_1, X_2, ..., X_n\right)$	¿ $\mathrm{IC}\left(\mu_1-\mu_2\right)$?	$m \rightarrow \left(Y_1, Y_2, ..., Y_m\right)$
$\downarrow$		$\downarrow$
$\overline{X}$	Muestras independientes	$\overline{Y}$

Recuerde que σ_1^2, σ_2^2 conocidas: $\quad \left(\overline{X}-\overline{Y}\right) \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right) \rightarrow$

(T. 2. 4. 1) $$\rightarrow \frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} = Z \sim N(0,1)$$

Una vez fijado el nivel de confianza $\beta = 1-\alpha$ establecemos que:

$$P(-Z_0 < Z < Z_0) = \beta = 1-\alpha$$



(Z_0 es el cuantil $1-\dfrac{\alpha}{2}$ de una $Z \sim N(0,1)$, por lo que $P(Z \le Z_0) = 1-\dfrac{\alpha}{2}$)

$$P\left(-Z_0 < \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n}+\dfrac{\sigma_2^2}{m}}} < Z_0 \right) = \beta = 1-\alpha$$

Se quiere construir un I. C. para $(\mu_1-\mu_2)$, por lo que en la expresión anterior habrá que realizar operaciones tendentes a que $(\mu_1-\mu_2)$ quede en la parte central de la desigualdad (las operaciones son análogas a las del apartado 2.1).

$$\boxed{P\left((\bar{X}-\bar{Y}) - Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} < (\mu_1-\mu_2) < (\bar{X}-\bar{Y}) + Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \right) = \beta = 1-\alpha} \ *$$

El centro del intervalo es $\left(\overline{X}-\overline{Y}\right)$ y su radio (cota superior o límite máximo para el error de estimación) es $r = Z_0\sqrt{\dfrac{\sigma_1^2}{n}+\dfrac{\sigma_2^2}{m}}$ (observe que es un intervalo simétrico respecto de $\left(\overline{X}-\overline{Y}\right)$).

$$\left(\overline{X}-\overline{Y}\right)$$

$$\left(\mu_1-\mu_2\right) \in \left(\qquad\qquad\qquad\right) \text{ con prob. } \beta = 1-\alpha$$

$$\left(\overline{X}-\overline{Y}\right)-Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \qquad Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \qquad Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \qquad \left(\overline{X}-\overline{Y}\right)+Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}$$

Error de estimación: $\left|\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)\right|$

La expresión * es equivalente a esta otra:

$$P\left(\left|\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)\right| < \underbrace{Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}_{=r}\right) = \beta = 1-\alpha$$

El error de estimación debe ser inferior a *r* (que actúa, por tanto, como cota superior o límite máximo para el error de estimación), con una probabilidad $\beta = 1-\alpha$.

El tamaño muestral necesario para que el error de estimación sea inferior a una cota dada de cuantía *r* (con un nivel de confianza $\beta = 1-\alpha$, también dado), se obtiene imponiendo que las dos muestras sean del mismo tamaño, y despejando el tamaño muestral de la expresión del radio, *r*:

$$n = m \rightarrow r = Z_0\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}\,;\; r = Z_0\sqrt{\frac{\sigma_1^2+\sigma_2^2}{n}}\,;\; r = Z_0\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sqrt{n}}\,;$$

$$\sqrt{n} = \frac{Z_0}{r}\sqrt{\sigma_1^2+\sigma_2^2}\,; \quad \boxed{n = \left(\frac{Z_0}{r}\right)^2\left(\sigma_1^2+\sigma_2^2\right)}$$

<u>Interpretación del resultado del I. C.</u> (aplicable a los cuatro casos del I. C. para la diferencia de medias: apartados 4.1., 4.2., 4.3. y 4.4.): cuando el I. C. obtenido incluye el 0, ello quiere decir que $\left(\mu_1-\mu_2\right)$ podría tomar el valor 0, con lo que no

existiría una diferencia significativa entre μ_1 y μ_2, con un nivel de confianza $\beta = 1-\alpha$.

4.2. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales desconocidas y muestras grandes (*n*+*m* > 30).

$$X \sim N(\mu_1;\sigma_1^2) \qquad \sigma_1^2,\ \sigma_2^2 \text{ desconocidas} \qquad Y \sim N(\mu_2;\sigma_2^2)$$

$$n \rightarrow (X_1, X_2, \ldots, X_n) \qquad ¿\mathrm{IC}(\mu_1-\mu_2)? \qquad m \rightarrow (Y_1, Y_2, \ldots, Y_m)$$

$$\bar{X} \quad S_n^2 \quad S_{n-1}^2 \qquad n+m>30 \qquad \bar{Y} \quad S_m^2 \quad S_{m-1}^2$$

(Muestras independientes)

Recuerde que:

σ_1^2, σ_2^2 desconocidas

Muestras grandes

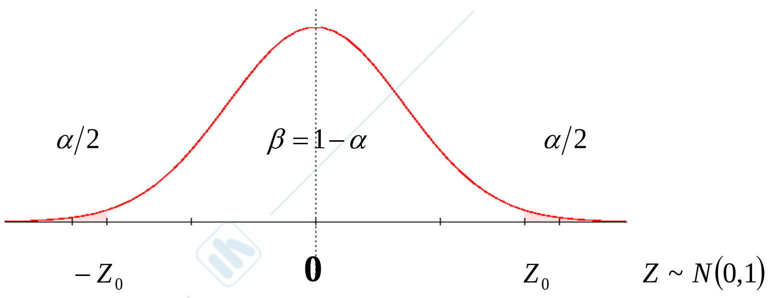
(T.2.4.2.)

$$\left.\begin{array}{l} (\bar{X}-\bar{Y}) \sim N\left(\mu_1-\mu_2,\ \dfrac{S_{n-1}^2}{n}+\dfrac{S_{m-1}^2}{m}\right) \\ (\bar{X}-\bar{Y}) \sim N\left(\mu_1-\mu_2,\ \dfrac{S_n^2}{n-1}+\dfrac{S_m^2}{m-1}\right) \end{array}\right\} \rightarrow$$

$$\rightarrow \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_{n-1}^2}{n}+\dfrac{S_{m-1}^2}{m}}} = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_n^2}{n-1}+\dfrac{S_m^2}{m-1}}} = Z \sim N(0,1)$$

Una vez fijado el nivel de confianza $\beta = 1-\alpha$ establecemos que:

$$P(-Z_0 < Z < Z_0) = \beta = 1-\alpha$$



(Z_0 es el cuantil $1-\dfrac{\alpha}{2}$ de una $Z \sim N(0,1)$, por lo que $P(Z \le Z_0) = 1-\dfrac{\alpha}{2}$)

$$P\left(-Z_0<\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{S_{n-1}^2}{n}+\frac{S_{m-1}^2}{m}}}<Z_0\right)=\beta=1-\alpha$$

(podría utilizarse la otra expresión del estadístico)

Tras realizar operaciones análogas a las del apartado 2.1., se obtiene la expresión del $\mathrm{IC}\left(\mu_1-\mu_2\right)$:

Si se utilizan S_{n-1}^2, S_{m-1}^2)

$$\boxed{P\left(\left(\overline{X}-\overline{Y}\right)-Z_0\sqrt{\frac{S_{n-1}^2}{n}+\frac{S_{m-1}^2}{m}}<\left(\mu_1-\mu_2\right)<\left(\overline{X}-\overline{Y}\right)+Z_0\sqrt{\frac{S_{n-1}^2}{n}+\frac{S_{m-1}^2}{m}}\right)=\beta=1-\alpha}$$

o bien, S_n^2, S_m^2)

$$\boxed{P\left(\left(\overline{X}-\overline{Y}\right)-Z_0\sqrt{\frac{S_n^2}{n-1}+\frac{S_m^2}{m-1}}<\left(\mu_1-\mu_2\right)<\left(\overline{X}-\overline{Y}\right)+Z_0\sqrt{\frac{S_n^2}{n-1}+\frac{S_m^2}{m-1}}\right)=\beta=1-\alpha}$$

Ambas expresiones dan finalmente el mismo resultado.

El centro del intervalo es $\left(\overline{X}-\overline{Y}\right)$ y su radio (cota superior o límite máximo para el error de estimación) es $r=Z_0\sqrt{\frac{S_{n-1}^2}{n}+\frac{S_{m-1}^2}{m}}$, o bien, $r=Z_0\sqrt{\frac{S_n^2}{n-1}+\frac{S_m^2}{m-1}}$.

Intervalo simétrico respecto de $\left(\overline{X}-\overline{Y}\right)$.

El tamaño muestral necesario para que el error de estimación sea inferior a una cota dada de cuantía r (con un nivel de confianza $\beta=1-\alpha$, también dado), se obtiene imponiendo que las dos muestras sean del mismo tamaño, y despejando el tamaño muestral de la expresión del radio, r:

$$n=m\rightarrow r=Z_0\sqrt{\frac{S_{n-1}^2}{n}+\frac{S_{m-1}^2}{n}}=Z_0\sqrt{\frac{S_{n-1}^2+S_{m-1}^2}{n}}\ ;\ r=Z_0\frac{\sqrt{S_{n-1}^2+S_{m-1}^2}}{\sqrt{n}}\ ;$$

$$\sqrt{n}=\frac{Z_0}{r}\sqrt{S_{n-1}^2+S_{m-1}^2}\ ;\qquad \boxed{n=\left(\frac{Z_0}{r}\right)^2\left(S_{n-1}^2+S_{m-1}^2\right)}$$

Si se utilizan las varianzas muestrales S_n^2, S_m^2:

$$n = m \rightarrow r = Z_0\sqrt{\frac{S_n^2}{n-1}+\frac{S_m^2}{n-1}} = Z_0\sqrt{\frac{S_n^2+S_m^2}{n-1}}\ ;\ r = Z_0\frac{\sqrt{S_n^2+S_m^2}}{\sqrt{n-1}}\ ;$$

$$\sqrt{n-1} = \frac{Z_0}{r}\sqrt{S_n^2+S_m^2}\ ;\ \boxed{n = \left(\frac{Z_0}{r}\right)^2\left(S_n^2+S_m^2\right)+1}$$

S_{n-1}^2 y S_{m-1}^2, o bien, S_n^2 y S_m^2, que aparecen en las anteriores expresiones del cálculo del tamaño muestral deben ser obtenidas mediante un muestreo realizado con anterioridad al actual, por lo que se requiere un muestreo bietápico.

4. 3. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales desconocidas, pero iguales, y muestras pequeñas ($n+m \leq 30$).

$X \sim N(\mu_1;\sigma_1^2)$
$\downarrow$
$n \rightarrow (X_1, X_2, \ldots, X_n)$
$\downarrow \searrow \searrow$
$\bar{X}$ S_n^2 S_{n-1}^2

σ_1^2, σ_2^2 des. pero $\sigma_1^2 = \sigma_2^2$

¿$\mathrm{IC}(\mu_1-\mu_2)$?

$n+m \leq 30$
(Muestras independientes)

$Y \sim N(\mu_2;\sigma_2^2)$
$\downarrow$
$m \rightarrow (Y_1, Y_2, \ldots, Y_m)$
$\downarrow \searrow \searrow$
$\bar{Y}$ S_m^2 S_{m-1}^2

Recuerde que:

σ_1^2, σ_2^2 desconocidas, pero $\sigma_1^2 = \sigma_2^2$

Muestras pequeñas
(T.2.4.3.)

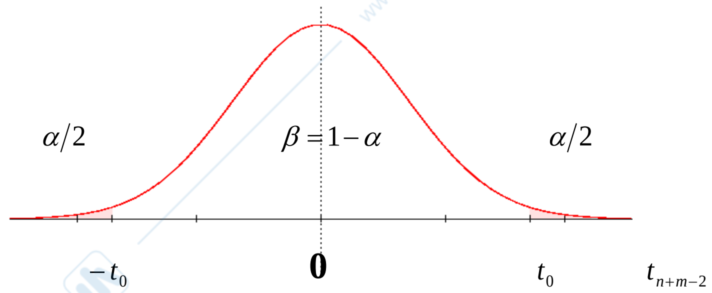
$$\left.\right\}\ \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_c\sqrt{\frac{1}{n}+\frac{1}{m}}} = t_{n+m-2}$$

Donde:

$$S_c = \sqrt{\frac{n S_n^{\,2} + m S_m^{\,2}}{n+m-2}} = \sqrt{\frac{(n-1)S_{n-1}^{\,2}+(m-1)S_{m-1}^{\,2}}{n+m-2}} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{j=1}^{m}(Y_j-\bar{Y})^2}{n+m-2}}$$

Una vez fijado el nivel de confianza $\beta = 1-\alpha$ establecemos que:

$$P(-t_0 < t_{n+m-2} < t_0) = \beta = 1-\alpha$$



(t_0 es el cuantil $1-\dfrac{\alpha}{2}$ de una t_{n+m-2}, por lo que $P\left(t_{n+m-2} \le t_0\right) = 1-\dfrac{\alpha}{2}$)

$$P\left(-t_0 < \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{S_c\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} < t_0 \right) = \beta = 1-\alpha$$

Se pretende construir un $\mathrm{IC}\left(\mu_1-\mu_2\right)$, por lo que haciendo operaciones análogas a las del apartado 2.1., se obtiene:

$$\boxed{P\left(\left(\bar{X}-\bar{Y}\right) - t_0\, S_c\sqrt{\frac{1}{n}+\frac{1}{m}} < \left(\mu_1-\mu_2\right) < \left(\bar{X}-\bar{Y}\right) + t_0\, S_c\sqrt{\frac{1}{n}+\frac{1}{m}} \right) = \beta = 1-\alpha}$$

Centro: $\left(\bar{X}-\bar{Y}\right)$. Radio: $r = t_0\, S_c\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}$. Intervalo simétrico respecto de $\left(\bar{X}-\bar{Y}\right)$.

4. 4. I. C. para la diferencia de medias poblacionales, con varianzas poblacionales desconocidas y desiguales, con muestras pequeñas (Aprox. Welch)

Recuerde que:

σ_1^2, σ_2^2 desconocidas y $\sigma_1^2 \ne \sigma_2^2$

Muestras pequeñas
(T.2.4.4.)

$$\left.\right\} \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{S_{n-1}^2}{n}+\dfrac{S_{m-1}^2}{m}}} = \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{S_n^2}{n-1}+\dfrac{S_m^2}{m-1}}} = t_v$$

Donde:

$$v = \frac{\left[\dfrac{\sigma_1^2/\sigma_2^2}{n}+\dfrac{1}{m}\right]^2}{\dfrac{\left(\sigma_1^2/\sigma_2^2\right)^2}{n^2\left(n-1\right)}+\dfrac{1}{m^2\left(m-1\right)}}$$

Generalmente el cociente σ_1^2 / σ_2^2 es desconocido, pudiendo sustituirse por el cociente S_{n-1}^2 / S_{m-1}^2 .

Siguiendo pasos análogos a los del epígrafe anterior se llega a:

Si se utilizan S_{n-1}^2, S_{m-1}^2)

$$\boxed{P\left(\left(\bar{X}-\bar{Y}\right) - t_0 \sqrt{\frac{S_{n-1}^2}{n} + \frac{S_{m-1}^2}{m}} < \left(\mu_1 - \mu_2\right) < \left(\bar{X}-\bar{Y}\right) + t_0 \sqrt{\frac{S_{n-1}^2}{n} + \frac{S_{m-1}^2}{m}} \right) = \beta = 1-\alpha}$$

o bien, S_n^2, S_m^2)

$$\boxed{P\left(\left(\bar{X}-\bar{Y}\right) - t_0 \sqrt{\frac{S_n^2}{n-1} + \frac{S_m^2}{m-1}} < \left(\mu_1 - \mu_2\right) < \left(\bar{X}-\bar{Y}\right) + t_0 \sqrt{\frac{S_n^2}{n-1} + \frac{S_m^2}{m-1}} \right) = \beta = 1-\alpha}$$

(t_0 es el cuantil $1-\dfrac{\alpha}{2}$ de una t_v , por lo que $P\left(t_v \le t_0\right) = 1-\dfrac{\alpha}{2}$).

Centro: $\left(\bar{X}-\bar{Y}\right)$. Radio: $r = t_0 \sqrt{\dfrac{S_{n-1}^2}{n} + \dfrac{S_{m-1}^2}{m}}$, o bien, $r = t_0 \sqrt{\dfrac{S_n^2}{n-1} + \dfrac{S_m^2}{m-1}}$

Intervalo simétrico respecto de $\left(\bar{X}-\bar{Y}\right)$.

El tamaño muestral necesario para que el error de estimación sea inferior a una cota dada de cuantía *r* (con un nivel de confianza $\beta = 1-\alpha$, también dado), se obtiene imponiendo que las dos muestras sean del mismo tamaño, y despejando el tamaño muestral de la expresión del radio, *r*:

<u>En el caso 4.3:</u>

($r = t_0\, S_c \sqrt{\frac{1}{n} + \frac{1}{m}}$)

$$\boxed{n = 2\left(\frac{\overset{Z_0}{t_0}\ S_c}{r}\right)^2}$$

<u>En el caso 4.4:</u>

$$\boxed{n = \left(\frac{\overset{Z_0}{t_0}}{r}\right)^2 \left(S_{n-1}^2 + S_{m-1}^2\right)}\text{; o bien: }\boxed{n = \left(\frac{\overset{Z_0}{t_0}}{r}\right)^2 \left(S_n^2 + S_m^2\right) + 1}$$

En las expresiones obtenidas se sustituye el cuantil t_0 por Z_0 .

S_{n-1}^2 y S_{m-1}^2, o bien, S_n^2 y S_m^2, que aparecen en las anteriores expresiones del cálculo del tamaño muestral deben ser obtenidas mediante un muestreo realizado con anterioridad al actual, por lo que se requiere un muestreo bietápico.

5. I. C. PARA EL COCIENTE DE VARIANZAS POBLACIONALES

La comparación de las varianzas entre dos poblaciones se realiza mediante cociente (en vez de hacerlo a través de la diferencia, como se ha visto para el caso de dos medias).

Caso A: μ_1, μ_2 desconocidas

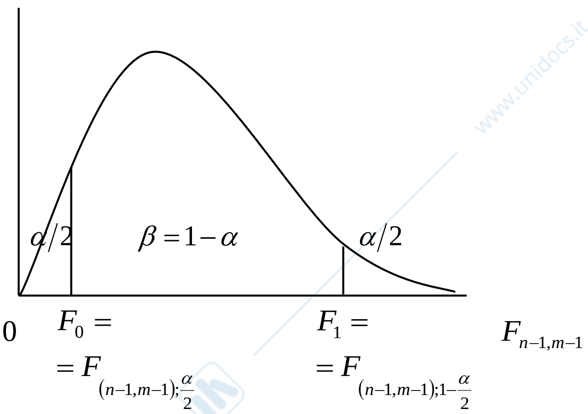
$X \sim N(\mu_1;\sigma_1^2)$	μ_1, μ_2 desconocidas	$Y \sim N(\mu_2;\sigma_2^2)$
$\downarrow$		$\downarrow$
$n \rightarrow (X_1, X_2, \ldots, X_n)$	¿ $\mathrm{IC}\left(\sigma_2^2/\sigma_1^2\right)$?	$m \rightarrow (Y_1, Y_2, \ldots, Y_m)$
$\downarrow \searrow \searrow$		$\downarrow \searrow \searrow$
$\bar{X}$ S_n^2 S_{n-1}^2	Muestras independientes	$\bar{Y}$ S_m^2 S_{m-1}^2

Recuerde que: $$\frac{S_{n-1}^2/\sigma_1^2}{S_{m-1}^2/\sigma_2^2} = F_{n-1,\, m-1} \; *$$

(T.2.4.5.)

Una vez fijado el nivel de confianza $\beta = 1-\alpha$ establecemos que:

$$P\left(F_0 < F_{n-1,\, m-1} < F_1\right) = \beta = 1-\alpha$$



$$P\left(F_{(n-1,\, m-1);\, \frac{\alpha}{2}} < F_{n-1,\, m-1} < F_{(n-1,\, m-1);1-\frac{\alpha}{2}}\right) = \beta = 1-\alpha$$

$F_1 = F_{(n-1,\, m-1);1-\frac{\alpha}{2}}$ puede encontrarse directamente en las tablas (cuando la probabilidad acumulada hasta F_1 sea 0'9, 0'95, 0'975, 0'99, 0'995). Pero el valor

$F_0 = F_{(n-1,m-1);\frac{\alpha}{2}}$ no puede encontrarse directamente en las tablas, sino que se obtiene indirectamente aplicando el teorema de inversión:

$$F_0 = \underbrace{F_{(n-1,m-1);\frac{\alpha}{2}}}_{\text{(no tabla)}} = \frac{1}{\underbrace{F_{(m-1,n-1);1-\frac{\alpha}{2}}}_{\text{(sí tabla)}}}$$

Así, el intervalo:

$$P\left(F_{(n-1,m-1);\frac{\alpha}{2}} < F_{n-1,m-1} < F_{(n-1,m-1);1-\frac{\alpha}{2}}\right) = \beta = 1-\alpha$$ se puede reescribir:

$$P\left(\frac{1}{F_{(m-1,n-1);1-\frac{\alpha}{2}}} < F_{n-1,m-1} < F_{(n-1,m-1);1-\frac{\alpha}{2}}\right) = \beta = 1-\alpha$$

A continuación, se sustituye la expresión * en el intervalo:

$$P\left(\frac{1}{F_{(m-1,n-1);1-\frac{\alpha}{2}}} < \frac{S_{n-1}^2/\sigma_1^2}{S_{m-1}^2/\sigma_2^2} < F_{(n-1,m-1);1-\frac{\alpha}{2}}\right) = \beta = 1-\alpha$$

Teniendo en cuenta que: $\dfrac{S_{n-1}^2/\sigma_1^2}{S_{m-1}^2/\sigma_2^2} = \dfrac{S_{n-1}^2}{\sigma_1^2} : \dfrac{S_{m-1}^2}{\sigma_2^2} = \dfrac{S_{n-1}^2\sigma_2^2}{\sigma_1^2 S_{m-1}^2} = \dfrac{S_{n-1}^2}{S_{m-1}^2}\cdot\dfrac{\sigma_2^2}{\sigma_1^2}$

$$P\left(\frac{1}{F_{(m-1,n-1);1-\frac{\alpha}{2}}} < \frac{S_{n-1}^2}{S_{m-1}^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} < F_{(n-1,m-1);1-\frac{\alpha}{2}}\right) = \beta = 1-\alpha$$

El objetivo es obtener el $\mathrm{IC}\left(\dfrac{\sigma_2^2}{\sigma_1^2}\right)$, por lo que el parámetro $\dfrac{\sigma_2^2}{\sigma_1^2}$ debe quedar sólo en la parte central de la desigualdad. Para ello, multiplicamos los tres miembros por $\dfrac{S_{m-1}^2}{S_{n-1}^2}$:

$S_{n-1}^2,\ S_{m-1}^2)$ $\boxed{P\left(\frac{1}{F_{(m-1,n-1);1-\frac{\alpha}{2}}}\frac{S_{m-1}^2}{S_{n-1}^2}<\frac{\sigma_2^2}{\sigma_1^2}<F_{(n-1,m-1);1-\frac{\alpha}{2}}\frac{S_{m-1}^2}{S_{n-1}^2}\right)=\beta=1-\alpha}$

Para obtener el intervalo en función de las varianzas muestrales ($S_n^2,\ S_m^2$) se hacen las siguientes sustituciones:

$$nS_n^2=(n-1)S_{n-1}^2 \rightarrow S_{n-1}^2=\frac{n}{n-1}S_n^2$$

$$mS_m^2=(m-1)S_{m-1}^2 \rightarrow S_{m-1}^2=\frac{m}{m-1}S_m^2$$

$S_n^2,\ S_m^2)$ $\boxed{P\left(\frac{1}{F_{(m-1,n-1);1-\frac{\alpha}{2}}}\frac{\frac{m}{m-1}S_m^2}{\frac{n}{n-1}S_n^2}<\frac{\sigma_2^2}{\sigma_1^2}<F_{(n-1,m-1);1-\frac{\alpha}{2}}\frac{\frac{m}{m-1}S_m^2}{\frac{n}{n-1}S_n^2}\right)=\beta=1-\alpha}$

<u>Interpretación del resultado del IC:</u> cuando el IC obtenido incluye la unidad, ello quiere decir que σ_2^2/σ_1^2 podría tomar el valor 1, con lo que no existiría una diferencia significativa entre σ_2^2 y σ_1^2, con un nivel de confianza β.

<u>Caso B:</u> $\mu_1,\ \mu_2$ conocidas

$X \sim N(\mu_1;\sigma_1^2)$	$\mu_1,\ \mu_2$ conocidas	$Y \sim N(\mu_2;\sigma_2^2)$
$\downarrow$		$\downarrow$
$n \rightarrow (X_1, X_2, \ldots, X_n)$	¿IC(σ_2^2/σ_1^2)?	$m \rightarrow (Y_1, Y_2, \ldots, Y_m)$
	Muestras independientes	

Se vio en T.2.2 y en el T.4.3 que cuando μ es conocida, se usa dicho parámetro en vez del estadístico $\overline{X}$, y, en tal caso el estadístico a utilizar:

$\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$ sigue una distribución χ_n^2

Por analogía:

$$\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{\sigma_1^2} \text{ sigue distribución } \chi_n^2 \qquad \frac{\sum_{j=1}^{m}(Y_j-\mu_2)^2}{\sigma_2^2} \text{ sigue distribución } \chi_m^2$$

Recuerde, además, que se vio en el Tema 1.7:

$$F_{n,m}=\frac{\frac{1}{n}\chi_n^2}{\frac{1}{m}\chi_m^2}$$

Teniendo esto en cuenta, el estadístico:

$$\frac{\frac{1}{n}\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{\sigma_1^2}}{\frac{1}{m}\frac{\sum_{j=1}^{m}(Y_j-\mu_2)^2}{\sigma_2^2}} \text{ sigue una distribución } F_{n,m}$$

Partiendo de ello, y siguiendo pasos análogos a los del Caso A (se propone hacer el desarrollo), se llega al IC:

$$\boxed{P\left(\frac{1}{F_{(m,n);1-\frac{\alpha}{2}}}\frac{\frac{\sum_{j=1}^{m}(Y_j-\mu_2)^2}{m}}{\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{n}}<\frac{\sigma_2^2}{\sigma_1^2}<F_{(n,m);1-\frac{\alpha}{2}}\frac{\frac{\sum_{j=1}^{m}(Y_j-\mu_2)^2}{m}}{\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{n}}\right)=\beta=1-\alpha}$$

6. I. C. PARA PROPORCIONES (POBLACIONES BINOMIALES)

6. 1. I. C. PARA LA PROPORCIÓN POBLACIONAL

Consideremos una población sobre la cual se estudia cierta característica, representada por la variable aleatoria X, de tal forma que cada elemento de la población, individualmente considerado, puede tener o no dicha característica. Desde este punto de vista, la variable aleatoria X podría conceptuarse como una binomial. Llamaremos:

p = proporción de elementos de la población que tienen la característica

$q = 1 - p$ = proporción de elementos de la población que no tienen la característica

Para construir un I. C. para p, extraemos una muestra aleatoria simple, a partir de la cual se pueden obtener las proporciones muestrales $\hat{p}$ y $\hat{q}$:

X (Binomial)

$\left.\begin{matrix} p \\ q \end{matrix}\right\}$ proporciones poblacionales

$n > 30$ IC (p)

$\hat{p}$ $\hat{q}$ (proporciones muestrales)

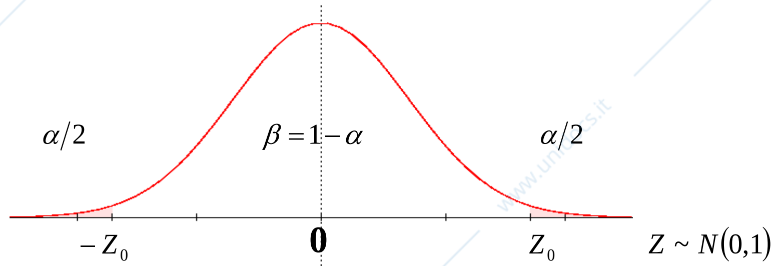
$\hat{p}$ = proporción de elementos de la muestra que tienen la característica

$\hat{q}$ = proporción de elementos de la muestra que no tienen la característica

Se vio en T.2.5.1 que si $n > 30$: $\hat{p} \sim N\left(p, \frac{pq}{n}\right) \rightarrow \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = Z \sim N(0,1)$

Una vez fijado el nivel de confianza $\beta = 1 - \alpha$ establecemos que:

$$P(-Z_0 < Z < Z_0) = \beta = 1 - \alpha$$



(Z_0 es el cuantil $1 - \frac{\alpha}{2}$ de una $Z \sim N(0,1)$, por lo que $P(Z \leq Z_0) = 1 - \frac{\alpha}{2}$)

se sustituye:

$$P\left(-Z_0 < \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} < Z_0\right) = \beta = 1 - \alpha$$

Se quiere construir un I. C. para p, por lo que, haciendo operaciones análogas a las del apartado 2.1., se llega a:

$$P\left(\hat{p} - Z_0\sqrt{\frac{pq}{n}} < p < \hat{p} + Z_0\sqrt{\frac{pq}{n}}\right) = \beta = 1 - \alpha$$

Pero este intervalo aún no se puede calcular porque p y q son desconocidos, por lo que no se puede calcular $\sqrt{\frac{pq}{n}}$. Éste se puede estimar mediante su correspondiente estimador muestral $\sqrt{\frac{\hat{p}\hat{q}}{n}}$, que sí se puede calcular: $\sqrt{\frac{pq}{n}} \xrightarrow{\text{estimación}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$. El IC quedaría así:

$$\boxed{P\left(\hat{p} - Z_0\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_0\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = \beta = 1 - \alpha}$$

El IC obtenido está centrado en $\hat{p}$ y tiene de radio $r = Z_0\sqrt{\frac{\hat{p}\hat{q}}{n}}$ (que es el límite máximo o cota superior para el error de estimación).

Para determinar el tamaño de la muestra que es necesario tomar para estimar p (con un límite máximo o cota superior para el error de estimación de cuantía r dado, y con un nivel de confianza β también dado), se despeja n de la expresión del radio, r:

dados r, β: $r = Z_0\sqrt{\frac{\hat{p}\hat{q}}{n}}$; $r = Z_0\frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}}$; $\sqrt{n} = \frac{Z_0}{r}\sqrt{\hat{p}\hat{q}}$; $\boxed{n = \left(\frac{Z_0}{r}\right)^2(\hat{p}\hat{q})}$

$\hat{p}$ y $\hat{q}$ son las proporciones de una muestra anterior a la que ahora se pretende extraer (muestreo bietápico). Si no se dispone de esta información muestral previa, nos situamos en el caso de la máxima incertidumbre, asignándole a $\hat{p}$ y $\hat{q}$ el valor 0'5: $\hat{p} = \hat{q} = 0'5$, sustituyéndolas en la ecuación de n:

$$\boxed{n = \left(\frac{Z_0}{r}\right)^2(0'25)}$$

Este tamaño muestral es más grande que el que se obtiene cuando hay información muestral previa sobre $\hat{p}$ y $\hat{q}$ (se obtiene un tamaño muestral denominado "conservador").

6. 2. I. C. PARA LA DIFERENCIA DE PROPORCIONES POBLACIONALES

Consideremos una población sobre la cual se estudia cierta característica, representada por la v. a. *X*, tal que cada elemento de la población, individualmente considerado, puede tener o no dicha característica.

Consideremos otra población sobre la cual se estudia cierta característica, representada por la v. a. *Y*, tal que cada elemento de la población, individualmente considerado, puede tener o no dicha característica.

X (Binomial)

p_1=proporción de elementos de la población 1 que tienen la característica

$q_1=1-p_1$=proporción de elementos de la población 1 que no tienen la característica

n

$\hat{p}_1 \quad \hat{q}_1 \; (\hat{q}_1 = 1 - \hat{p}_1)$

Y (Binomial)

p_2=proporción de elementos de la población 2 que tienen la característica

$q_2=1-p_2$=proporción de elementos de la población 2 que no tienen la característica

m

$\hat{p}_2 \quad \hat{q}_2 \; (\hat{q}_2 = 1 - \hat{p}_2)$

Muestras independientes, con $n + m > 30$

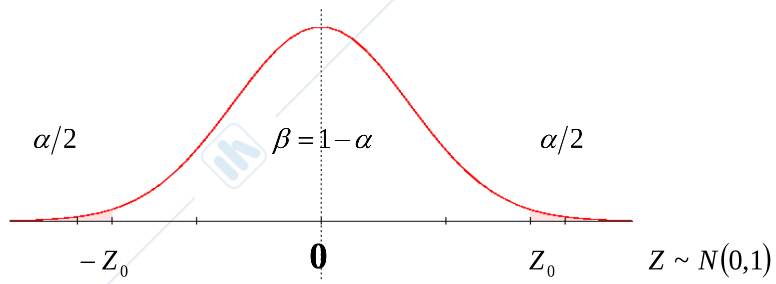
¿IC (p_1-p_2)?

Recuerde que si $n+m>30$: $\quad (\hat{p}_1 - \hat{p}_2) \sim N\left(p_1 - p_2, \dfrac{p_1q_1}{n} + \dfrac{p_2q_2}{m}\right) \rightarrow$

(T. 2. 5. 2.) $\quad \rightarrow \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1q_1}{n} + \dfrac{p_2q_2}{m}}} = Z \sim N(0,1)$

Una vez fijado el nivel de confianza $\beta = 1 - \alpha$ establecemos que:

$$P(-Z_0 < Z < Z_0) = \beta = 1 - \alpha$$



(Z_0 es el cuantil $1-\dfrac{\alpha}{2}$ de una $Z \sim N(0,1)$, por lo que $P(Z \le Z_0) = 1 - \dfrac{\alpha}{2}$)

se sustituye:

$$P\left(-Z_0 < \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}} < Z_0 \right) = \beta = 1 - \alpha$$

Se quiere construir un I. C. para $(p_1 - p_2)$, por lo que, haciendo operaciones:

$$P\left((\hat{p}_1 - \hat{p}_2) - Z_0 \sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + Z_0 \sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}} \right) = \beta = 1 - \alpha$$

En la expresión anterior, generalmente $\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}$ será desconocida, pero puede ser estimada mediante su correspondiente estimador muestral, $\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}}$. Por lo que la expresión del IC quedaría así:

$$\boxed{P\left((\hat{p}_1 - \hat{p}_2) - Z_0 \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + Z_0 \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}} \right) = \beta = 1 - \alpha}$$

Interpretación del resultado del IC: cuando el IC construido contiene el valor 0, ello querría decir que $(p_1 - p_2)$ podría tomar el valor 0, por lo que no existiría una diferencia significativa entre p_1 y p_2, con un nivel de confianza β.

El IC obtenido está centrado en $(\hat{p}_1 - \hat{p}_2)$ y tiene de radio $r = Z_0 \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}}$ (cota superior o límite máximo para el error de estimación).

Para determinar los tamaños que han de tener ambas muestras para que el error de estimación sea inferior a la cota dada, *r*, y para un nivel de confianza β también prefijado, se ha de imponer $n=m$ y despejar el tamaño muestral de la expresión de *r*:

dados *r*, β:

$$r = Z_0 \sqrt{\frac{\hat{p}_1 \hat{q}_1 + \hat{p}_2 \hat{q}_2}{n}}; \qquad \sqrt{n} = \frac{Z_0}{r} \sqrt{\hat{p}_1 \hat{q}_1 + \hat{p}_2 \hat{q}_2}; \qquad \boxed{n = \left(\frac{Z_0}{r} \right)^2 (\hat{p}_1 \hat{q}_1 + \hat{p}_2 \hat{q}_2)}$$

Análogamente a lo comentado en el epígrafe anterior, para obtener las proporciones muestrales $\hat{p}_1$, $\hat{q}_1$, $\hat{p}_2$, $\hat{q}_2$ hay dos soluciones:

1ª) Si se conocen gracias a un muestreo anterior, se utiliza dicha información previa (muestreo bietápico).

2ª) Si no se cuenta con información anterior, se opta por asignar a las proporciones muestrales el valor 0'5, por lo que:

$$n = \left(\frac{Z_0}{r}\right)^2 (0'5 \cdot 0'5 + 0'5 \cdot 0'5); \qquad \boxed{n = \left(\frac{Z_0}{r}\right)^2 (0'5)}$$

Así se obtiene un tamaño muestral mayor que el que se obtendría si se tuviera alguna información sobre dichas proporciones muestrales (se obtendría un tamaño muestral denominado "conservador").