ESTAD_T24. ESTIMACIÓN POR INTERVALOS.

MÉTODOS de CONSTRUCCIÓN:
- Cantidad pivotal
- Método general de Neyman

INTERVALOS de CONFIANZA en POBL. NORMALES.
- Media
- Varianza
- Diferencia de medias
- Cociente de varianzas

REGIONES de CONFIANZA

---

0. INTRODUCCIÓN.

Inferencia → Estimación → { Puntual / Intervalos

0. INFERENCIA. INTRODUCCIÓN

~~1. ESTIMACIÓN PUNTUAL I.~~

Recordemos brevemente que la Inferencia Estadística consiste en ~~generalizar~~ (sacar conclusiones) sobre la poblac. que nos interesa estudiar a partir ~~de~~ la información que nos proporciona una muestra aleatoria basándonos en la Teoría de la Probabilidad.

Si estamos interesados en ~~en~~ estudiar el valor de una característica poblacional θ, la Inferencia que podemos utilizar es:
- Estimación
- Contrastación

La Estimación consiste en dar un valor aproximado del parám. poblacional a partir de la información ~~proporcionada~~ muestral.
La Contrastación consiste es formular una conjetura (hipótesis) sobre el valor del parámetro poblacional y utilizar la inform. muestral para aceptar ó rechazar dicha hipótesis.

En el caso de la estimación, se puede aproximar:
- Estim. puntual → valor concreto
- Estim. por intervalos de confianza → intervalo

En ambos métodos se utiliza un estadístico (función real de la muestra) para estimar.
La ventaja de la estimación puntual es que da un valor concreto del estimador que puede sustituirse directamente en el parámetro, pero sólo cuenta con las propiedades como parámetro.
La ventaja de la estimación por intervalos es que el intervalo va acompañado de un grado de confianza de que el verdadero valor del parámetro se encuentre dentro del intervalo, pero su inconveniente es que ofrece infinitas soluciones.

1. ESTIMACIÓN POR INTERVALOS

Como alternativa a la estimación puntual, que no informa sobre la proximidad de la estimación del verdadero valor del parámetro poblacional, se utiliza la estimación por intervalos que acompaña la estimación puntual de un intervalo donde confiamos que se encuentre incluido el verdadero valor del parámetro θ.

$$I.C.(\theta) \longrightarrow [\underline{\theta}(X), \overline{\theta}(X)] \text{ donde } \begin{cases} \underline{\theta}(X) \text{ límite inferior} \\ \overline{\theta}(X) \text{ límite superior} \end{cases}$$

de modo que:

$$P\left(\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right) = 1-\alpha \equiv \text{nivel de confianza} \begin{cases} \text{tanto por 1} \\ \% \end{cases}$$

- El $IC(\theta)$ es un intervalo aleatorio (antes de particularizarse para una muestra concreta), que depende de la muestra X,
- θ no es aleatorio, es un parámetro desconocido.

La expresión $P(\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)) = 1-\alpha$ hay que interpretarla de la siguiente manera:

→ Como probabilidad antes de particularizar ~~una~~ la muestra, porque son v.aleatorias.

→ Como confianza después de obtener la muestra, porque:
- ya no hay aleatoriedad, $P=0$ ó $P=1$
- no tenemos manera de saberlo (θ descon.)

En el ejemplo de Arnáiz:
- Estudiante se sabe 99 de 100 temas, pero no recuerda el nº del que no se sabe. El examen es sacar bola y cantar.
Antes de sacar bola, P(Aprobar) = 0'99 → probab.
Después de sacar bola, Confianza → lo suerte está echada, pero entre la extracción y la revisión del profesor tiene una confianza del 99%.

2_ MÉTODOS de CONSTRUCCIÓN: pivotal y general.

Los métodos buscan obtener los límites del IC.

2.1_ MT. PIVOTE

Primero definimos la cantidad pivotal y después exponemos el método.

Para una población ξ, con función de probabilidad $f(x;\theta)$ donde θ es un parámetro real desconocido, extraemos una m.a.s. (n)

Una cantidad pivotal o pivote $T(X_1 \dots X_n, \theta)$ es una función muestral que depende de θ, $\begin{cases} \bar{X} \text{ fija, sólo depende de } \theta \\ \theta \text{ fija, depende de } \bar{X} \end{cases}$ pero su distrib. de probabilidad no depende de θ

Ejemplo:

$\xi \to N(\mu, \sigma)$ σ conocido.

$\hat{\mu} = \bar{X} \to N(\mu, \sigma/\sqrt{n}) \Longrightarrow \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ cantidad pivotal.

Teorema: Si la cantidad pivotal $T(\bar{X}, \theta)$ es función monótona de θ, entonces es posible determinar el IC para θ.

Se pueden elegir $K_1(\alpha)$ y $K_2(\alpha)$ para un nivel de confianza $1-\alpha$, pertenecientes al campo de variación de $T(\bar{X}, \theta)$, tales que:

$$P\big(K_1(\alpha) \leq \theta \leq K_2(\alpha)\big) = 1 - \alpha \quad , \forall \theta \in \Theta$$

Si $T(\bar{X}, \theta)$ es monótona en θ, se pueden resolver las ecuaciones:

$$\left.\begin{array}{l} T(\bar{X}, \theta) = K_1(\alpha) = \underline{\theta}(\bar{X}) \\ T(\bar{X}, \theta) = K_2(\alpha) = \overline{\theta}(\bar{X}) \end{array}\right\} \Rightarrow IC(\theta) = [\underline{\theta}(\bar{X}), \overline{\theta}(\bar{X})].$$

2.2 _ MT. GENERAL de Neyman

Ventajas: No exige la propiedad del pivote (distrib. probab. del pivote indep de θ).

Puede utilizar una función muestral cuya distrib. dependa de θ.

~~Inconvenientes~~:

Sea una población ξ, con función de probabilidad $f(x,\theta)$, se quiere construir un IC para θ.

Se toma un estimador de θ, $\hat{\theta}$ con función de probab. $g(\hat{\theta},\theta)$.

Para un nivel de confianza $1-\alpha$ dado, se determinan los extremos de un intervalo $K_1(\alpha,\theta)$ y $K_2(\alpha,\theta)$ tq:

$$P\left[K_1(\alpha,\theta) \le \theta \le K_2(\alpha,\theta)\right] = 1-\alpha$$

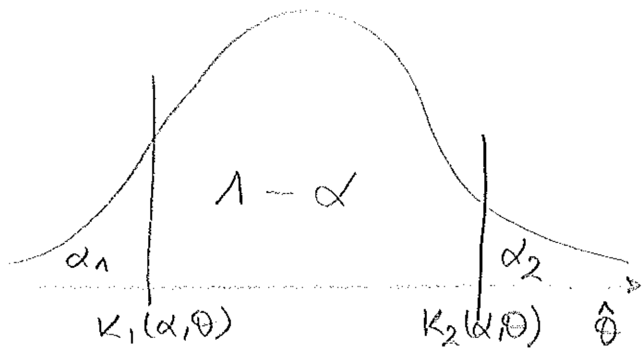
En una población continua tendremos:

$$\int_{K_1(\alpha,\theta)}^{K_2(\alpha,\theta)} g(\hat{\theta},\theta)\, d\hat{\theta} = 1-\alpha$$

o bien

$$\int_{-\infty}^{K_1(\alpha,\theta)} g(\hat{\theta},\theta)\, d\hat{\theta} = \alpha_1$$

$$\int_{K_2(\alpha,\theta)}^{+\infty} g(\hat{\theta},\theta)\, d\hat{\theta} = \alpha_2$$



El reparto de la probab. entre las colas indica la longitud del intervalo obtenido.

Al resolver las integrales anteriores se obtienen los límites del intervalo $K_1(\alpha,\theta)$ y $K_2(\alpha,\theta)$ tales que

$P[K_1(\alpha,\theta) \le \theta \le K_2(\alpha,\theta)] = 1-\alpha$

de donde se deducen las ecuaciones integrales:

$K_2(\alpha,\theta) = \hat{\theta}_0 \Rightarrow \underline{\theta}(X) = K_2^{-1}(\alpha,\hat{\theta}_0)$

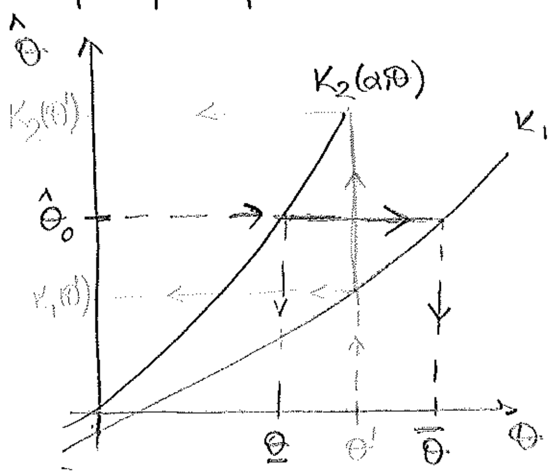
$K_1(\alpha,\theta) = \hat{\theta}_0 \Rightarrow \overline{\theta}(X) = K_1^{-1}(\alpha,\hat{\theta}_0)$

$\Big\}$ IC $\rightarrow [\underline{\theta}, \overline{\theta}]$.

Gráficamente:

Representamos $K_1(\alpha,\theta)$ y $K_2(\alpha,\theta)$ funciones de θ

Sp. que para una mas(n) X el valor de $\hat{\theta}$ es $\hat{\theta}_0$



$\hat{\theta} = \hat{\theta}_0 \longrightarrow \{ \text{I.C.} \rightarrow [\underline{\theta}, \overline{\theta}] \}$.

$\theta = \theta' \Rightarrow P[K_1(\theta') \le \hat{\theta}_0 \le K_2(\theta')] = 1-\alpha$.

Sp. que la muestra procede de una población en la que el verdadero valor del parámetro es $\theta = \theta'$, entonces la probabilidad de que la estimación $\hat{\theta}_0$, para esa muestra, esté comprendida entre $K_1(\theta')$ y $K_2(\theta')$ será $1-\alpha$.
Gráficamente, si se corta la horizontal y la vertical el intervalo $[\underline{\theta}, \overline{\theta}]$ incluye al verdadero valor del parámetro.

De aquí se deducen las ecuaciones integrales:

$K_1(\alpha,\overline{\theta}) = \hat{\theta}_0$

$K_2(\alpha,\underline{\theta}) = \hat{\theta}_0$

$\Big\}$ y teniendo en cuenta

$\int_{-\infty}^{\hat{\theta}_0} g(\hat{\theta},\theta)\,d\hat{\theta} = \alpha_1$

$\int_{\hat{\theta}_0}^{+\infty} g(\hat{\theta},\theta)\,d\hat{\theta} = \alpha_2$

los resultados serán los extremos del intervalo de confianza $[\underline{\theta}, \overline{\theta}]$ con un nivel de confianza $1-\alpha$.

3_ INTERVALOS de CONFIANZA en POBL. NORMALES

(1) I.C. para la media, μ

Sea una población $\xi \to N(\mu, \sigma)$, σ conocida;

$\hat{\mu} = \bar{x} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ de donde se deduce el pivote:

$$T(X, \mu) = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$$

$$IC(\alpha) \longrightarrow P[K_1 \le T(X, \mu) \le K_2] = 1-\alpha$$

$$1-\alpha = P\left(K_1 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le K_2\right) = P\left(K_1 \cdot \frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le K_2 \cdot \frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(-\bar{X} - K_1 \cdot \frac{\sigma}{\sqrt{n}} \le -\mu \le -\bar{X} - K_2 \cdot \frac{\sigma}{\sqrt{n}}\right) = P\left(\bar{X} - K_2 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} - K_1 \frac{\sigma}{\sqrt{n}}\right)$$

luego $IC(\alpha) \longrightarrow \left[\bar{X} - K_2 \frac{\sigma}{\sqrt{n}}, \bar{X} - K_1 \cdot \frac{\sigma}{\sqrt{n}}\right]$

cuya longitud es $D = \left(\bar{X} - K_1 \frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - K_2 \frac{\sigma}{\sqrt{n}}\right) = (K_2 - K_1) \frac{\sigma}{\sqrt{n}}$

Se plantea minimizar la distancia del intervalo sujeto a la condición del nivel de confianza:

$$\min_{K_1, K_2} \ (K_2 - K_1)\frac{\sigma}{\sqrt{n}}$$

$$\text{s.a. } P[K_1 \le N(0,1) \le K_2] = 1-\alpha$$

$$\min \phi = (K_2 - K_1)\frac{\sigma}{\sqrt{n}} + \lambda \left[\int_{K_1}^{K_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du - (1-\alpha)\right] = 0.$$

$$\left.\begin{aligned} \frac{\partial \phi}{\partial K_2} &= \frac{\sigma}{\sqrt{n}} + \lambda \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}K_2^2} = 0 \\ \frac{\partial \phi}{\partial K_1} &= -\frac{\sigma}{\sqrt{n}} - \lambda \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}K_1^2} = 0 \end{aligned}\right\} \longrightarrow e^{-\frac{1}{2}K_2^2} = e^{-\frac{1}{2}K_1^2} \longrightarrow K_2^2 = K_1^2$$

$$\frac{\partial \phi}{\partial \lambda} = \int_{K_1}^{K_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du - (1-\alpha) = 0$$

$$K_2^2 = K_1^2 \iff \begin{cases} K_1 = K_2 \Rightarrow D = \frac{\sigma}{\sqrt{n}}(K_2 - K_1) = 0 \quad !! \\ K_1 = -K_2 \Rightarrow \text{Intervalo simétrico.} \end{cases}$$

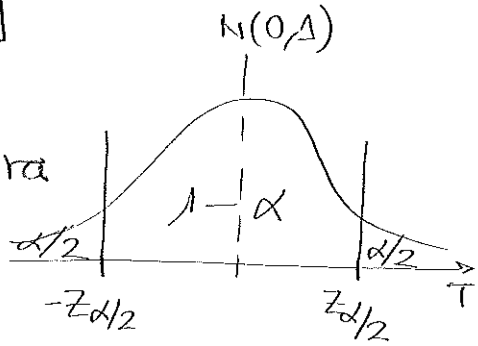
luego

$$IC(\mu_\sigma)\Big|_\alpha \longrightarrow \left[\bar{X} - K\cdot\frac{\sigma}{\sqrt{n}} \, , \, \bar{X} + K\cdot\frac{\sigma}{\sqrt{n}}\right]$$

donde K depende del nivel de confianza

$$K = z_{\alpha/2} \to P(N(0,1) > z_{\alpha/2}) = \frac{\alpha}{2}$$

↳ tabla $N(0,1)$.



Sea una población $\xi \to N(\mu, \sigma)$ σ desconocida:

$$\hat{\mu} = \bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\left.\begin{array}{l} \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1) \\ \dfrac{nS^2}{\sigma^2} \to \chi^2_{n-1} \end{array}\right\} \; t_{n-1} = \frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}} = \frac{\dfrac{\bar{X}-\mu}{\cancel{\sigma}/\sqrt{n}}}{\sqrt{\dfrac{nS^2}{\cancel{\sigma^2}}\Big/ n-1}} =$$

$$T(X, \mu) = \frac{\sqrt{n-1}\,(\bar{X}-\mu)}{S} \quad \text{pivote}$$

→ depende de μ

→ su distrib (t_{n-1}) no depende de μ.

Por la simetría de la distrib. t_{n-1}, se llega a un intervalo simétrico como intervalo de longitud mínima. Siguiendo el método del pivote:

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

$$= P\left[-t\cdot\frac{S}{\sqrt{n-1}} \le \bar{X}-\mu \le t\cdot\frac{S}{\sqrt{n-1}}\right] = \ldots =$$

$$= P\left[\bar{X} - t\cdot\frac{S}{\sqrt{n-1}} \le \mu \le \bar{X} + t\cdot\frac{S}{\sqrt{n-1}}\right] \Rightarrow IC(\mu_S)\Big|_\alpha \to \left[\bar{X} \pm t\cdot\frac{S}{\sqrt{n-1}}\right]$$

(2) I.C. para la varianza, σ^2

Sea una población $\xi \longrightarrow N(\mu, \sigma)$, se trata de construir un I.C. para la varianza poblacional.

Martín-Pliego no distingue el caso de media conocida del de media desconocida, y utiliza como estadístico la varianza muestral.

Casas sí hace distinción, y en el caso de media descon. utiliza la cuasivarianza muestral.

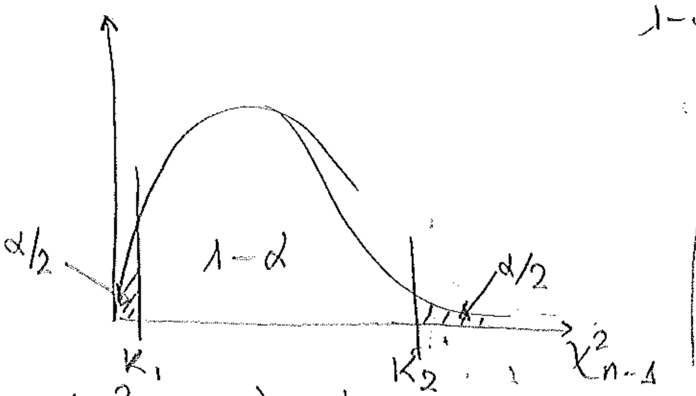
Pero todas las cantidades pivotales se distribuyen con arreglo a una distrib χ^2, por lo que el razonamiento es análogo.

Lo vemos:

varianza muestral

M-P $\longrightarrow$ Cantidad pivotal: $\dfrac{nS^2}{\sigma^2} \longrightarrow \chi^2_{n-1}$

Como χ^2 no es simétrica y su ~~distribución~~ forma depende de los grados de libertad, se elige el intervalo asignando la misma probab. a la izqda y * a la derecha del mismo.



$$1-\alpha = P\left[K_1 \leq \chi^2_{n-1} \leq K_2\right] =$$
$$= P\left[K_1 \leq \frac{nS^2}{\sigma^2} \leq K_2\right] =$$
$$= P\left[\frac{K_1}{nS^2} \leq \frac{1}{\sigma^2} \leq \frac{K_2}{nS^2}\right] =$$
$$= P\left[\frac{nS^2}{K_2} \leq \sigma^2 \leq \frac{nS^2}{K_1}\right]$$
$$\longrightarrow IC(\sigma^2)\Big|_{\alpha} \longrightarrow \left[\frac{nS^2}{K_2}, \frac{nS^2}{K_1}\right].$$

$$\left.\begin{array}{l} P(\chi^2_{n-1} > K_2) = \frac{\alpha}{2} \Rightarrow K_2 \\ P(\chi^2_{n-1} > K_1) = 1 - \frac{\alpha}{2} \Rightarrow K_1 \end{array}\right\} \text{tablas } \chi^2$$

$$* \; S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i}{n} \longrightarrow \chi^2_{n-1}$$ porque tiene una ligadura $(\bar{x})$

$\Longrightarrow$ Datos

ESTAD _ T24 ⑨

Fotocopias Casas pág. 242, 290–295

(3) I.C. para la diferencia de medias poblacionales (INDEP)

a) Con varianzas conocidas:

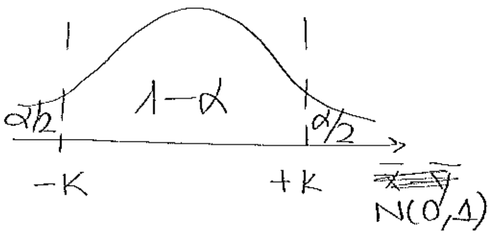
$\xi_1 \longrightarrow N(\mu_1, \sigma_1) \xrightarrow{mas(n)} X = \{x_1 \dots x_n\}, \bar{x}$

$\xi_2 \longrightarrow N(\mu_2, \sigma_2) \xrightarrow{mas(m)} Y = \{y_1 \dots y_m\}, \bar{y}$

$$\left.\begin{array}{l} \bar{x} \longrightarrow N\left(\mu_1, \dfrac{\sigma_1}{\sqrt{n}}\right) \\ \bar{y} \longrightarrow N\left(\mu_2, \dfrac{\sigma_2}{\sqrt{m}}\right) \end{array}\right\} (\bar{x}-\bar{y}) \longrightarrow N\left(\mu_1-\mu_2, \sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)$$

$$T(X, Y, \mu_1, \mu_2) = \frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n}+\dfrac{\sigma_2^2}{m}}} \longrightarrow N(0,1)$$

de la que se deduce el intervalo de confianza simétrico:



$$IC \longrightarrow \left[(\bar{x}-\bar{y})-K\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}, (\bar{x}-\bar{y})+K\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right]$$

b) Con varianzas desconocidas (por simplicidad, op $\sigma_1=\sigma_2$).

$\xi_1 \longrightarrow N(\mu_1, \sigma_1) \xrightarrow{mas(n)} X = \{x_1 \dots x_n\}\ \bar{x}, S_x^2$

$\xi_2 \longrightarrow N(\mu_2, \sigma_2) \xrightarrow{mas(m)} Y = \{y_1 \dots y_m\}\ \bar{y}, S_y^2$.

$$\left.\begin{array}{l} \dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{m\sigma^2+n\sigma^2}{n\cdot m}}} \longrightarrow N(0,1) \\ \\ \dfrac{(nS_x^2+mS_y^2)}{\sigma^2} \longrightarrow \chi^2_{n+m-2} \end{array}\right\} t_{n+m-2} = \frac{\dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{(m+n)\cancel{\sigma^2}}{n\cdot m}}}}{\sqrt{\dfrac{nS_x^2+mS_y^2}{\cancel{\sigma^2}(n+m-2)}}}$$

cuyo intervalo de confianza para $1-\alpha$ será

$$I.C. \rightarrow \left[(\bar{x}-\bar{y}) \pm t \cdot \sqrt{\frac{m+n}{n\cdot m}} \cdot \sqrt{\frac{nS_x^2+mS_y^2}{n+m-2}}\right]$$

ESTAD_T24 DUDA

(4) I.C. para el cociente de varianzas

$\mathcal{S}_1 \to N(\mu_1, \sigma_1) \xrightarrow{\text{mas (n)}} X=\{x_1 \dots x_n\}\ \bar{x},\ S_x^2$

$\mathcal{S}_2 \to N(\mu_2, \sigma_2) \xrightarrow{\text{mas (m)}} Y=\{y_1 \dots y_m\}\ \bar{y},\ S_y^2$

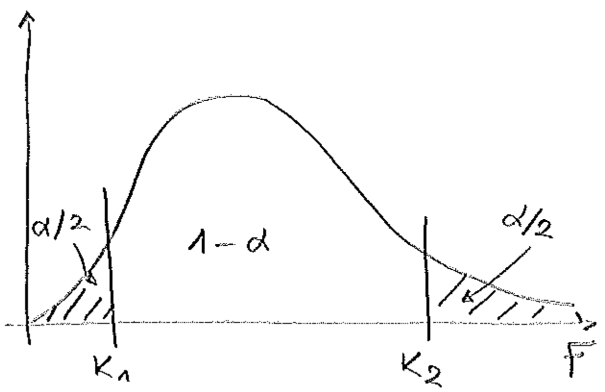
INDEP

MP→ Por el lema de Fisher-Cochran:

$$\left.\begin{array}{l} \dfrac{nS_x^2}{\sigma_1^2} \longrightarrow \chi^2_{n-1} \\ \dfrac{mS_y^2}{\sigma_2^2} \longrightarrow \chi^2_{m-1} \end{array}\right\} \quad T\left(X,Y,\frac{\sigma_1^2}{\sigma_2^2}\right) = \frac{\dfrac{nS_x^2}{\sigma_1^2}}{\dfrac{mS_y^2}{\sigma_2^2}} \longrightarrow F_{n-1,m-1}$$

¿no le falta dividir por los grados de libertad?

Se utiliza el criterio simplificador, próximo al de longitud mínima de dejar la misma probab. en cada lado del intervalo.



$P(F_{n-1,m-1} \geq k_2) = \alpha/2 \Rightarrow k_2$

$P(F_{n-1,m-1} \geq k_1) = 1-\alpha/2 \Rightarrow k_1$

$$1-\alpha = P[K_1 \leq F_{n-1,m-1} \leq K_2] =$$

$$= P\left[K_1 \leq \frac{\dfrac{nS_x^2}{\sigma_1^2}}{\dfrac{mS_y^2}{\sigma_2^2}} \leq K_2\right] =$$

$$= P\left[K_1 \cdot \frac{mS_y^2}{nS_x^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq K_2 \cdot \frac{mS_y^2}{nS_x^2}\right] =$$

$$= P\left[\frac{nS_x^2}{mS_y^2} \cdot \frac{1}{K_2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{nS_x^2}{mS_y^2} \cdot \frac{1}{K_1}\right]$$

Casas sí divide por los grados de libertad, que se anulan al utilizar la cuasivarianza:

μ desc. $\longrightarrow T\left(X,Y,\frac{\sigma_1^2}{\sigma_2^2}\right) = \dfrac{\dfrac{\cancel{(n-1)}S_{1x}^2}{\sigma_1^2} \Big/ \cancel{n-1}}{\dfrac{\cancel{(m-1)}S_{1y}^2}{\sigma_2^2} \Big/ \cancel{m-1}} = \dfrac{S_{1x}^2}{S_{1y}^2} \cdot \dfrac{\sigma_2^2}{\sigma_1^2} \longrightarrow F_{n-1,m-1}$

I.C. $\longrightarrow \left[\dfrac{S_{1x}^2}{S_{1y}^2} \cdot \dfrac{1}{k_2} \leq \dfrac{\sigma_1^2}{\sigma_2^2} \leq \dfrac{S_{1x}^2}{S_{1y}^2} \cdot \dfrac{1}{k_1}\right]$

⟹ mirar atrás

4_REGIONES de CONFIANZA

En poblaciones multiparamétricas donde su distrib. de probabilidad $f(x, \theta_1, \dots, \theta_K)$ depende de más de un parámetro es posible determinar una región de confianza para más de un parámetro de dimensión $\mathbb{R}^K$ (K = nº parám. que nos interesan).

$$P[(\theta_1, \dots, \theta_K) \in R(X)] = 1-\alpha$$

No necesariamente la región de confianza se tiene que referir a todos los parámetros, nos puede interesar un subgrupo de parámetros, considerando al resto parámetros perturbadores o accesorios.

Para construir una región de confianza no se puede, en general, partir de los IC unidimensionales, porque puede ocurrir que los estadísticos utilizados NO sean indep.

La cantidad pivotal dependerá de cada caso, el planteamiento es el mismo.

Ejemplo: Región de confianza para μ y σ^2 de una $N(\mu, \sigma^2)$.

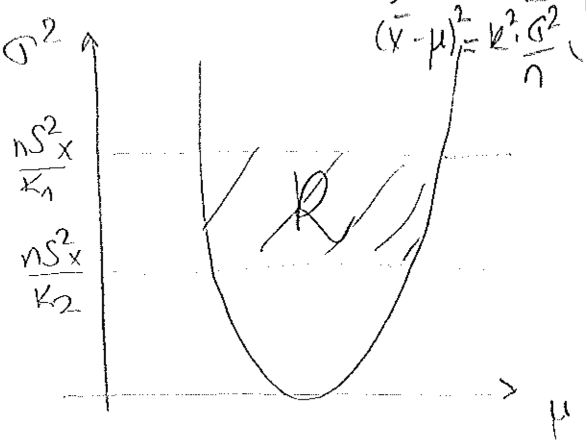
Por el lema de Fisher-Cochran $\bar{x}$ y nS_x^2 son indep, luego

$$\left.\begin{array}{l} \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1) \\ \dfrac{nS_x^2}{\sigma^2} \longrightarrow \chi^2_{n-1} \end{array}\right\} \quad 1-\alpha = P[(\mu, \sigma^2) \in R(\bar{x}, nS_x^2)] =$$

$$= P[\mu \in R_1(\bar{x})] \cdot P[\sigma^2 \in R_2(nS_x^2)] =$$

$$= \underbrace{P\left[\bar{x} - K\cdot\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + K\cdot\frac{\sigma}{\sqrt{n}}\right]}_{\sqrt{1-\alpha}} \cdot \underbrace{P\left(\frac{nS_x^2}{K_2} \le \sigma^2 \le \frac{nS_x^2}{K_1}\right)}_{\sqrt{1-\alpha}}$$

$$(\bar{x}-\mu)^2 = K^2 \cdot \frac{\sigma^2}{n}$$



R no es la de área mínima, pero sí la más sencilla de calcular.

$$P\left[k_1(\alpha,\theta) \le \theta \le k_2(\alpha,\theta)\right] = 1-\alpha$$

de donde se deducen las <u>ecuaciones integrales</u>:

$$k_1(\alpha,\theta) = \hat{\theta}$$
$$k_2(\alpha,\theta) = \hat{\theta}$$

~~de uno~~

de modo que si $k_1(\alpha,\theta)$ y $k_2(\alpha,\theta)$ son funciones monótonas de θ, permiten despejar el parámetro θ y obtener los límites del intervalo.

OJO → Si la población es discreta, a lo mejor no se puede obtener un $IC(\alpha)$ sino aproximado.

2_3. <u>Intervalos de confianza de longitud mínima</u>

Tal y como se ha planteado el mt. general, existen infinitos α_1, α_2 que verifican $\alpha_1+\alpha_2=\alpha$, lo que lleva a obtener distintos intervalos con el mismo nivel de confianza.

Para determinar el intervalo de longitud mínima, se utiliza el mt. de optimización condicionada de los multiplicadores de Lagrange.

* No siempre es factible obtener un intervalo de nivel de confianza $1-\alpha$ de longitud mínima.

Una solución alternativa es minimizar la longitud esperada

$$\min \; E\left[\overline{\theta}(X) - \underline{\theta}(X)\right]$$
$$\text{s.a.} \; P(\theta \in IC) = 1-\alpha$$

→ Solución: el criterio más habitualmente empleado es repartir el complementario del nivel de confianza en dos partes iguales que, coincide muchas veces con el intervalo de longitud mínima.

4.9. CUADRO RESUMEN DE LOS INTERVALOS DE CONFIANZA

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
$N(\mu, \sigma)$	μ μ	$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ $Z \to N(0,1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu, \sigma)$	μ, σ μ	$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	n pequeña $\left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ $t \to t_{n-1} \quad , \quad P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$
Desconocida	μ, σ μ	$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	n grande $\left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ $Z \to N(0, 1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu\ \sigma)$	μ, σ σ	$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$	n pequeña $\left[\frac{(n-1)s^2}{\chi^2_{n-1,\ 1-\alpha/2}} \leqslant \sigma^2 \leqslant \frac{(n-1)s^2}{\chi^2_{n-1,\ \alpha/2}}\right]$ $P[\chi^2_{n-1} \leqslant \chi^2_{n-1, 1-\alpha/2}] = 1 - \frac{\alpha}{2} \quad , \quad P[\chi^2_{n-1} \leqslant \chi^2_{n-1,\ \alpha/2}] = \frac{\alpha}{2}$
$N(\mu, \sigma)$	μ, σ σ	$\hat{\sigma}^2 = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	n grande $\left[\frac{s^2}{1 + z_{\alpha/2}\sqrt{\frac{2}{n-1}}} \leqslant \sigma^2 \leqslant \frac{s^2}{1 - z_{\alpha/2}\sqrt{\frac{2}{n-1}}}\right]$ $Z_1 \to N(0,1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu, \sigma)$	σ σ	$\hat{\sigma}^2 = s^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$	n pequeña $\left[\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\chi^2_{n,\ 1-\alpha/2}} \leqslant \sigma^2 \leqslant \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\chi^2_{n,\ \alpha/2}}\right]$ $P[\chi^2_n \leqslant \chi^2_{n,\ 1-\alpha/2}] = 1 - \frac{\alpha}{2} \quad ; \quad P[\chi^2_n \leqslant \chi^2_{n,\ \alpha/2}] = \frac{\alpha}{2}$
$N(\mu, \sigma)$	σ σ	$\hat{\sigma}^2 = s^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$	n grande $\left[\frac{s^{*2}}{1 + z_{\alpha/2}\sqrt{\frac{2}{n}}} \leqslant \sigma^2 \leqslant \frac{s^{*2}}{1 - z_{\alpha/2}\sqrt{\frac{2}{n}}}\right]$ $Z \to N(0,1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x}\sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y}\sum_{i=1}^{n_y} y_i$	$\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$ $\left. \leqslant (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right]$ $Z \to N(0, 1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y σ_x, σ_y $\sigma_x = \sigma_y$ $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x}\sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y}\sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{n_y} (y_i - \bar{y})^2$	n_x, n_y pequeñas $\left[(\bar{x} - \bar{y}) - t_{\alpha/2}\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}\sqrt{\frac{n_x + n_y}{n_x n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$ $\left. \leqslant (\bar{x} - \bar{y}) + t_{\alpha/2}\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}\sqrt{\frac{n_x + n_y}{n_x n_y}}\right]$ $t \to t_{n_x + n_y - 2} \quad , \quad P[t_{n_x + n_y - 2} > t_{\alpha/2}] = \frac{\alpha}{2}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
Desconocidas	μ_x, μ_y σ_x, σ_y $\sigma_x = \sigma_y$ $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x}\sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y}\sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{n_y} (y_i - \bar{y})^2$	n_x, n_y grandes $\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}\sqrt{\frac{n_x + n_y}{n_x n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$ $\left. \leqslant (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}\sqrt{\frac{n_x + n_y}{n_x n_y}}\right]$ $Z \to N(0, 1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y σ_x, σ_y $\sigma_x \neq \sigma_y$ $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x}\sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y}\sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{n_y} (y_i - \bar{y})^2$	n_x, n_y pequeñas $\left[(\bar{x} - \bar{y}) - t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant (\bar{x} - \bar{y}) + t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right]$ $t \to t_\nu$ (t-Student) $\quad \nu = \dfrac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\dfrac{\left(\frac{s_x^2}{n_x}\right)^2}{n_x - 1} + \dfrac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y - 1}}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
Desconocidas	μ_x, μ_y σ_x, σ_y $\sigma_x \neq \sigma_y$ $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$	n_x, n_y grandes $\left[(\bar{x} - \bar{y}) - z_{\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$ $\left. \leqslant (\bar{x} - \bar{y}) + z_{\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right]$ $Z \to N(0, 1)$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$ Apareadas	$\mu_D = \mu_x - \mu_y$ μ_D	$\hat{\bar{D}} = \bar{d} = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)$ $\hat{\sigma}_D^2 = s_d^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (d_i - \bar{d})^2$	n pequeño $\left[\bar{d} - t_{\alpha/2} \frac{s_d}{\sqrt{n}} \leqslant \mu_D \leqslant \bar{d} + t_{\alpha/2} \frac{s_d}{\sqrt{n}}\right]$ $t \to t_{n-1} \quad , \quad P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y σ_x, σ_y $\frac{\sigma_x^2}{\sigma_y^2}$	$\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$	$\left[\frac{s_x^2}{s_y^2} \cdot \frac{1}{F_{n_x-1,\ n_y-1,\ 1-\alpha/2}} \leqslant \frac{\sigma_x^2}{\sigma_y^2} \leqslant \frac{s_x^2}{s_y^2} \cdot \frac{1}{F_{n_x-1,\ n_y-1,\ \alpha/2}}\right]$ $F_{n_x-1,\ n_y-1,\ 1-\alpha/2} = \frac{1}{F_{n_x-1,\ n_y-1,\ \alpha/2}}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	σ_x, σ_y $\frac{\sigma_x^2}{\sigma_y^2}$	$\hat{\sigma}_x^2 = s_x^{*2} = \frac{1}{n_x} \sum_{i=1}^{n_x} (x_i - \mu_x)^2$ $\hat{\sigma}_y^2 = s_y^{*2} = \frac{1}{n_y} \sum_{i=1}^{n_y} (y_i - \mu_y)^2$	$\left[\frac{s_x^{*2}}{s_y^{*2}} \cdot \frac{1}{F_{n_x,\ n_y,\ 1-\alpha/2}} \leqslant \frac{\sigma_x^2}{\sigma_y^2} \leqslant \frac{s_x^{*2}}{s_y^{*2}} \cdot \frac{1}{F_{n_x,\ n_y,\ \alpha/2}}\right]$
$B(1, p)$	p p	$\hat{p} = \frac{x}{n} = \frac{\text{n.º de éxitos en } n \text{ pruebas}}{\text{n.º de pruebas}}$	n pequeño Gráficos: Tabla A.13 n grande $\left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \leqslant p \leqslant \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}\right]$ $Z \to N(0, 1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$B(1, p_X)$ $B(1, p_Y)$	p_X, p_Y $p_X - p_Y$	$\hat{p}_x = \frac{x}{n}$ $\hat{p}_y = \frac{y}{n}$	$\left[(\hat{p}_x - \hat{p}_y) - z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1 - \hat{p}_x)}{n_x} + \frac{\hat{p}_y(1 - \hat{p}_y)}{n_y}} \leqslant p_X - p_Y \leqslant \right.$ $\left. \leqslant (\hat{p}_x - \hat{p}_y) + z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1 - \hat{p}_x)}{n_x} + \frac{\hat{p}_y(1 - \hat{p}_y)}{n_y}}\right]$ $Z \to N(0, 1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$