Dpto. de Economía Cuantitativa
Universidad Complutense de Madrid

Introducción a la Econometría

Tema 4 — Estimación y muestreo

Marcos Bujosa y
Gustavo A. Marrero

Material de apoyo para el curso *Introducción a la Econometría* de la licenciatura en Economía de la Universidad Complutense de Madrid.

Actualizado el: 10 de abril de 2007
Versión 4.0

Puede encontrar la última versión de este material en:

http://www.ucm.es/info/ecocuan/mbb/index.html#ietria

Índice

Este es un material de apoyo a las clases. En ningún caso sustituye a los libros de texto que figuran en el programa de la asignatura; textos que el alumno debe estudiar para afrontar el examen final con ciertas garantías de éxito.

El programa se cubre con los siguientes capítulos de libro de texto Novales (1997)[1]:

Capítulos 1 a 3: Estos temas han sido cubiertos en asignaturas anteriores, y debido a su bajo nivel de complejidad no se verán en clase (aunque forman parte del programa).

Capítulos 4 a 6: Estos temas han sido cubiertos en las asignaturas Estadística I y II. Se realizará un breve repaso en clase (una semana o semana y media como máximo), asumiendo que el alumno es capaz de preparar por su cuenta esta parte.

Capítulos 7 y 8: completos

Capítulo 9: secciones 9.4 a 9.6

Capítulos 10 y 12: completos

[1] Otros excelentes manuales en castellano son Peña (2001), Peña (2002) y Peña y Romo (1997).

1. Introducción

$$\begin{aligned} S \rightarrow & \quad \mathbb{R}_X \\ (\mathfrak{B},\ P(\cdot)) \rightarrow & \quad \{f_X(x;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \end{aligned}$$

Llamamos Soporte al conjunto $\mathbb{R}_X$ de posibles valores de la v.a.

Llamamos Modelo de Probabilidad a la familia o colección de funciones de densidad

$$\Phi = \{f_X(x;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta},\ x \in \mathbb{R}_X\}$$

Llamamos Espacio Paramétrico al conjunto $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ de posibles valores de los parámetros de las funciones de densidad.

1. Postulamos a priori una familia de densidades como mecanismo estocástico subyacente.

1. Postulamos a priori una familia de densidades como mecanismo estocástico subyacente.
   a) debemos elegir aquel Modelo de Probabilidad más adecuado a los datos
      - histograma $\to f_X(x; \boldsymbol{\theta})$
      - datos pueden restringir $\mathbb{R}_X$ y espacio paramétrico, $\boldsymbol{\Theta}$

1. Postulamos a priori una familia de densidades como mecanismo estocástico subyacente.

   a) debemos elegir aquel Modelo de Probabilidad más adecuado a los datos

      - histograma $\rightarrow f_X(x;\boldsymbol{\theta})$
      - datos pueden restringir $\mathbb{R}_X$ y espacio paramétrico, $\boldsymbol{\Theta}$

   b) Elegimos una familia (no una función en particular)

$$\Phi = \{f_X(x;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta},\ x \in \mathbb{R}_X\}$$

   donde los parámetros son desconocidos.

1. Postulamos a priori una familia de densidades como mecanismo estocástico subyacente.
   a) debemos elegir aquel Modelo de Probabilidad más adecuado a los datos
      - histograma $\rightarrow f_X(x; \boldsymbol{\theta})$
      - datos pueden restringir $\mathbb{R}_X$ y espacio paramétrico, $\boldsymbol{\Theta}$

   b) Elegimos una familia (no una función en particular)

$$\Phi = \{f_X(x; \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta},\ x \in \mathbb{R}_X\}$$

   donde los parámetros son desconocidos.

2. Inferencia estadística $\Rightarrow$ modelo de probabilidad (Tema 4 y siguientes)

Estudiar propiedades y/o relaciones estocásticas de poblaciones

$$\begin{aligned} X &: \quad \text{Renta familia Madrileña} \\ Y &: \quad \text{Renta familia Barcelonesa} \end{aligned}$$

Estudiar propiedades y/o relaciones estocásticas de poblaciones

$$X: \quad \text{Renta familia Madrileña}$$
$$Y: \quad \text{Renta familia Barcelonesa}$$

Podemos preguntarnos por

$$\mathrm{E}(Y) = 10000?; \quad \mathrm{E}(Y) \in (8000, 15000)?;$$
$$\mathrm{E}_{Y|X}(Y|\, 5000) > \mathrm{E}(Y)?;$$
$$\mathrm{Var}(Y) > \mathrm{Var}(X); \quad P(Y > X); \quad f_Y(y) \sim N(\mu_Y, \sigma_Y^2)$$

Estudiar propiedades y/o relaciones estocásticas de poblaciones

$$X: \quad \text{Renta familia Madrileña}$$
$$Y: \quad \text{Renta familia Barcelonesa}$$

Podemos preguntarnos por

$$\mathrm{E}\,(Y) = 10000?; \quad \mathrm{E}\,(Y) \in (8000, 15000)?;$$
$$\mathrm{E}_{Y|X}\,(Y|\,5000) > \mathrm{E}\,(Y)?;$$
$$\mathrm{Var}\,(Y) > \mathrm{Var}\,(X)\,; \quad P\,(Y > X)\,; \quad f_Y\,(y) \sim N(\mu_Y, \sigma_Y^2)$$

Conociendo $f_{XY}\,(x, y; \boldsymbol{\theta})$ respuesta inmediata.

Estudiar propiedades y/o relaciones estocásticas de poblaciones

$$\begin{aligned} X: &\quad \text{Renta familia Madrileña} \\ Y: &\quad \text{Renta familia Barcelonesa} \end{aligned}$$

Podemos preguntarnos por

$$\mathrm{E}\,(Y) = 10000?; \quad \mathrm{E}\,(Y) \in (8000, 15000)?;$$
$$\mathrm{E}_{Y|X}\,(Y|\,5000) > \mathrm{E}\,(Y)?;$$
$$\mathrm{Var}\,(Y) > \mathrm{Var}\,(X)\,; \quad P\,(Y > X)\,; \quad f_Y\,(y) \sim N(\mu_Y, \sigma_Y^2)$$

Conociendo $f_{XY}\,(x, y; \boldsymbol{\theta})$ respuesta inmediata.

Pero...

A veces sólo conocemos $f_{XY}\,(x, y)$ pero no $\boldsymbol{\theta}$ (Inferenc. paramétrica)

Estudiar propiedades y/o relaciones estocásticas de poblaciones

$$X: \quad \text{Renta familia Madrileña}$$
$$Y: \quad \text{Renta familia Barcelonesa}$$

Podemos preguntarnos por

$$\mathrm{E}\,(Y) = 10000?; \quad \mathrm{E}\,(Y) \in (8000, 15000)?;$$
$$\mathrm{E}_{Y|X}\,(Y|\,5000) > \mathrm{E}\,(Y)?;$$
$$\mathrm{Var}\,(Y) > \mathrm{Var}\,(X)\,; \quad P\,(Y > X)\,; \quad f_Y\,(y) \sim N(\mu_Y, \sigma_Y^2)$$

Conociendo $f_{XY}\,(x, y; \boldsymbol{\theta})$ respuesta inmediata.

Pero...

A veces sólo conocemos $f_{XY}\,(x, y)$ pero no $\boldsymbol{\theta}$ (Inferenc. paramétrica)

A veces ni siquiera conocemos $f_{XY}\,(x, y)$ (Inferenc. no paramétrica)

- Estimación puntual *vs* Intervalo de Confianza

- Contraste de hipótesis ("Tª de la decisión")
  - Paramétricos: $\boldsymbol{\theta}$ en $f_X(x;\boldsymbol{\theta})$
  - No paramétricos: (Distribución, Homogeneidad, Independencia)

2. Inferencia paramétrica

Momentos Teóricos (o momentos de una distribución): valor esperado de funciones $g(X)$

$$\mathrm{E}\left(g(X)\right) = \int_{\mathbb{R}_X} g(x) \cdot f_X\left(x; \boldsymbol{\theta}\right) dx$$

Momentos Teóricos (o momentos de una distribución): valor esperado de funciones $g(X)$

$$\mathrm{E}\left(g(X)\right) = \int_{\mathbb{R}_X} g(x) \cdot f_X(x;\boldsymbol{\theta})\, dx$$

por tanto son funciones $h(\boldsymbol{\theta}; x)$ de los parámetros

$$\mathrm{E}\left(g(X)\right) = h(\boldsymbol{\theta}; x)$$

Momentos Teóricos *(o momentos de una distribución):* valor esperado de funciones $g(X)$

$$\mathrm{E}\left(g(X)\right) = \int_{\mathbb{R}_X} g(x) \cdot f_X\left(x; \boldsymbol{\theta}\right) dx$$

por tanto son funciones $h(\boldsymbol{\theta}; x)$ de los parámetros

$$\mathrm{E}\left(g(X)\right) = h(\boldsymbol{\theta}; x)$$

Inferencia: sobre valor de $\boldsymbol{\theta}$ mediante *momentos muestrales* (dada $\boldsymbol{x}$).

Llamamos modelo muestral de tamaño n, $\boldsymbol{X}$, a

$$\boldsymbol{X} \equiv (X_1, X_2, \ldots, X_n)$$

y llamamos muestra a una realización de $\boldsymbol{X}$

$$\boldsymbol{x} \equiv (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n_{\boldsymbol{x}}.$$

Supondremos los datos observados como realizaciones del modelo muestral (i.e., una muestra).

Modelo estadístico

1. Modelo probabilístico

$$f_{X_i}(x_i;\boldsymbol{\theta})$$

2. Modelo muestral

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$$

Modelo estadístico

1. Modelo probabilístico

$$f_{X_i}(x_i; \boldsymbol{\theta})$$

2. Modelo muestral

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$$

$\Longrightarrow$

Distribución conjunta de $\boldsymbol{X}$

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\theta})$$

Modelo estadístico

1. Modelo probabilístico

$$f_{X_i}(x_i;\boldsymbol{\theta})$$

2. Modelo muestral

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$$

$\Longrightarrow$

Distribución conjunta de $\boldsymbol{X}$

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\theta})$$

Datos observados (muestra)

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

Realización de $\boldsymbol{X}$
(un punto del soporte $\mathbb{R}^n_{\boldsymbol{X}}$)

Modelo estadístico

1. Modelo probabilístico

$$f_{X_i}(x_i;\boldsymbol{\theta})$$

2. Modelo muestral

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$$

$\Rightarrow$

Distribución conjunta de $\boldsymbol{X}$

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\theta})$$

$\Downarrow$

Datos observados (muestra)

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

Realización de $\boldsymbol{X}$

(un punto del soporte $\mathbb{R}^n_{\boldsymbol{X}}$)

$\Rightarrow$

Inferencia estadística

- Estimación
- Contrastación
- Predicción

Llamamos *estadístico* a una función del modelo muestral

$$g(X_1, X_2, \ldots, X_n) \equiv g(\boldsymbol{X});$$

es por tanto *v.a.*

Llamamos *estadístico* a una función del modelo muestral

$$g(X_1, X_2, \ldots, X_n) \equiv g(\boldsymbol{X});$$

es por tanto *v.a.*

Sus momentos respecto al origen son

$$\mathrm{E}\left(\left[g(\boldsymbol{X})\right]^r\right) = \int \cdots \int_{\mathbb{R}^n_{\boldsymbol{X}}} \left[g(\boldsymbol{x})\right]^r f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\theta}) dx_1 \cdots dx_n$$

donde $\boldsymbol{x} \equiv (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n_{\boldsymbol{X}}$.

El *muestreo aleatorio simple (m.a.s.)* supone que X_i son idéntica e independientemente distribuidas

$$X_i \sim \text{iid}; \qquad i = 1, \ldots, n.$$

El *muestreo aleatorio simple (m.a.s.)* supone que X_i son idéntica e independientemente distribuidas

$$X_i \sim \text{iid}; \qquad i = 1, \ldots, n.$$

Por tanto, $f_{X_1}(x) = f_{X_2}(x) \cdots = f_{X_n}(x) \qquad = f_X(x)$;
es decir, cada x_i proviene de una misma población.

El *muestreo aleatorio simple (m.a.s.)* supone que X_i son idéntica e independientemente distribuidas

$$X_i \sim \mathrm{iid}; \qquad i = 1, \ldots, n.$$

Por tanto, $f_{X_1}(x) = f_{X_2}(x) \cdots = f_{X_n}(x) \qquad = f_X(x);$
es decir, cada x_i proviene de una misma población.
Además (por la independencia)

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x) = (f_X(x))^n$$

y

$$M_{X_1,\ldots,X_n} = \prod_{i=1}^{n} M_{X_i} = (M_X)^n$$

3. Estimación puntual y por intervalos (repaso)

Supongamos una muestra aleatoria simple (*m.a.s.*)

$$\boldsymbol{x} \equiv (x_1, x_2, \ldots, x_n);$$

realización del muestreo aleatorio simple (modelo muestral)

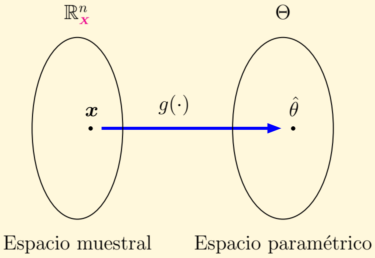
$$\boldsymbol{X} \equiv (X_1, X_2, \ldots, X_n) \text{ con } f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta}).$$

Buscamos un valor para el vector $\boldsymbol{\theta}$ que "*de algún modo*" es más acorde con la muestra $\boldsymbol{x}$.

x es *una* de las posibles realizaciones de X.

$\boldsymbol{x}$ es *una* de las posibles realizaciones de $\boldsymbol{X}$.

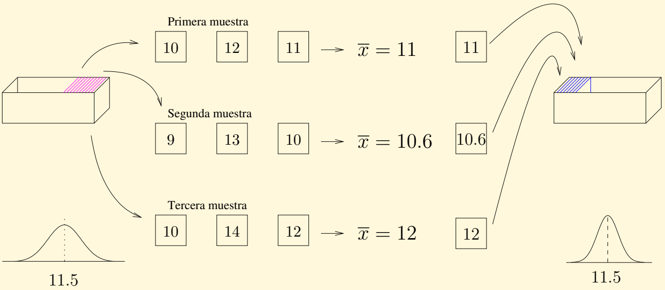
Buscamos una regla que elija un valor de $\boldsymbol{\theta}$ para cada $\boldsymbol{x}$.



estimación: $\widehat{\theta} = g(x_1, x_2 \ldots, x_n) = g(\boldsymbol{x})$

- (*Resumen de algún aspecto de la información muestral*)

estimador: $\widehat{\boldsymbol{\theta}} = g(X_1, X_2, \ldots, X_n) = g(\boldsymbol{X})$



- Modelo muestral: (X_1, X_2, X_3)
- Estimador: $\widehat{\theta} = (X_1 + X_2 + X_3)/3$
- Primera muestra: $\boldsymbol{x} = (10, 12, 11)$
- Estimación con primera muestra: $\widehat{\theta} = 11$
  (es un valor puntual, una realización de $\widehat{\theta}$)

Boxmodels (Sumas y medias)

Existen infinidad de posibles estimadores.

- problema de la elección (*Insesgadez, eficiencia*).
  - Máxima verosimilitud
  - Mínimos cuadrados
  - Método de los momentos
  - Otros.

Existen infinidad de posibles estimadores.

- problema de la elección (*Insesgadez, eficiencia*).
  - Máxima verosimilitud
  - Mínimos cuadrados
  - Método de los momentos
  - Otros.

Estimación puntual es *tan sólo* una realización de $\widehat{\theta}$.
Estimación puntual es *poco* informativa.

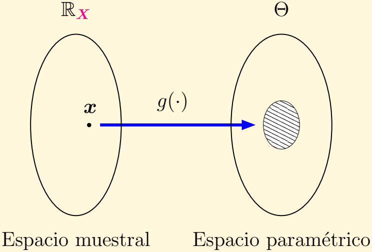
Existen infinidad de posibles estimadores.

- problema de la elección (*Insesgadez, eficiencia*).
  - Máxima verosimilitud
  - Mínimos cuadrados
  - Método de los momentos
  - Otros.

Estimación puntual es *tan sólo* una realización de $\widehat{\theta}$.
Estimación puntual es *poco* informativa.

Podemos decir más si conocemos la distribución de $\widehat{\theta}$.

Buscamos una regla que elija, para cada $\boldsymbol{x}$, un conjunto de valores que pueda contener el verdadero valor de $\boldsymbol{\theta}$ con cierta probabilidad.



Deseamos que:

- el conjunto sea pequeño (precisión)
- la probabilidad de que *el verdadero* $\boldsymbol{\theta}$ pertenezca al conjunto sea grande (confianza)

$$\left[h(\widehat{\theta_1}), h(\widehat{\theta_2})\right] = g(\boldsymbol{X}_{(n)})$$

de tal manera que

$$P\left(h(\widehat{\theta_1}) \leq \theta \leq h(\widehat{\theta_2})\right) = (1-\alpha)$$

donde

- $(1-\alpha)$ es el nivel de confianza.

Buscamos que $(1-\alpha)$ sea grande; y $(\widehat{\theta_2} - \widehat{\theta_1})$ sea mínimo.

(Generalmente se fija $(1-\alpha)$ y se busca que $(\widehat{\theta_2} - \widehat{\theta_1})$ sea mínimo.)

Necesitamos la distribución de $g(\boldsymbol{X}_{(n)})$

4. Distribuciones asociadas al muestreo (repaso)

4.1. Distribución de la media muestral

Estimador de la esperanza cuando la varianza es conocida

Sea $X \sim \mathrm{N}\left(\mu\,,\,\sigma^2\right)$ con σ^2 conocida.

Emplearemos como estimador de $\mathrm{E}\left(X\right)$ la media aritmética

$$g(X_1, \ldots, X_n) = \widehat{\overline{x}} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

donde

$$\widehat{\overline{x}} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

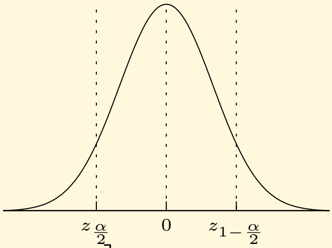
o tipificando

$$\frac{\widehat{\overline{x}} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Dado un nivel de confianza calculamos el intervalo

Puesto que $\frac{\hat{\bar{x}}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

$$P\left(a \leq \frac{\hat{\bar{x}}-\mu}{\sigma/\sqrt{n}} \leq b\right) = 1-\alpha$$



$$\mu \in \left[\hat{\bar{x}} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]; \quad \text{con probabilidad } 1-\alpha.$$

$$X_i : \text{Renta familia Madrileña } i$$
$$n = 16; \quad \sum x_i = 2536; \quad X_i \sim N(\mu, 250);$$
$$(1-\alpha) = 0.95 \implies z_{0.975} = 1.96$$

Entonces

$$\left[\frac{2536}{16} \pm \frac{\sqrt{250}}{\sqrt{16}} 1.96\right] = [151,025;\ 165,975]$$

$$X_i : \text{Renta familia Madrileña } i$$
$$n = 16; \quad \sum x_i = 2536; \quad X_i \sim N(\mu, 250);$$
$$(1-\alpha) = 0.95 \implies z_{0.975} = 1.96$$

Entonces

$$\left[\frac{2536}{16} \pm \frac{\sqrt{250}}{\sqrt{16}} 1.96\right] = [151,025;\ 165,975]$$

Cálculo con muestras A, B, C, ...

$$\widehat{\bar{x}}^{A} \quad = 150,5 \quad \rightarrow \quad [143.025;\ 157.975]$$

$$X_i : \text{Renta familia Madrileña } i$$
$$n = 16; \quad \sum x_i = 2536; \quad X_i \sim N(\mu, 250);$$
$$(1-\alpha) = 0.95 \implies z_{0.975} = 1.96$$

Entonces

$$\left[\frac{2536}{16} \pm \frac{\sqrt{250}}{\sqrt{16}} 1.96\right] = [151,025;\ 165,975]$$

Cálculo con muestras A, B, C, ...

$$\widehat{\overline{x}}^A = 150,5 \quad \rightarrow \quad [143.025;\ 157.975]$$
$$\widehat{\overline{x}}^B = 164 \quad \rightarrow \quad [156.525;\ 171.475]$$

$$X_i : \text{Renta familia Madrileña } i$$
$$n = 16; \quad \sum x_i = 2536; \quad X_i \sim N(\mu, 250);$$
$$(1-\alpha) = 0.95 \implies z_{0.975} = 1.96$$

Entonces

$$\left[\frac{2536}{16} \pm \frac{\sqrt{250}}{\sqrt{16}} 1.96\right] = [151,025;\ 165,975]$$

Cálculo con muestras A, B, C, ...

$$\begin{array}{lcl} \widehat{\overline{x}}^A = 150,5 & \rightarrow & [143.025;\ 157.975] \\ \widehat{\overline{x}}^B = 164 & \rightarrow & [156.525;\ 171.475] \\ \widehat{\overline{x}}^C = 155 & \rightarrow & [147.525;\ 162.475] \\ \ldots & \ldots & \ldots \end{array}$$

$$X_i : \text{Renta familia Madrileña } i$$
$$n = 16; \quad \sum x_i = 2536; \quad X_i \sim N(\mu, 250);$$
$$(1-\alpha) = 0.95 \implies z_{0.975} = 1.96$$

Entonces

$$\left[\frac{2536}{16} \pm \frac{\sqrt{250}}{\sqrt{16}} 1.96\right] = [151,025;\ 165,975]$$

Cálculo con muestras A, B, C, ...

$$\begin{array}{llll}
\widehat{\overline{x}}^A & = 150,5 & \rightarrow & [143.025;\ 157.975] \\
\widehat{\overline{x}}^B & = 164 & \rightarrow & [156.525;\ 171.475] \\
\widehat{\overline{x}}^C & = 155 & \rightarrow & [147.525;\ 162.475] \\
\ldots & & \ldots & \ldots
\end{array}$$

"verdadero valor" es $\mu = 155,2$ (pero es DESCONOCIDO!!!)

Boxmodels (Intervalos de confianza)

En economía generalmente se dispone de una única muestra (no hay laboratorio)

En economía generalmente se dispone de una única muestra (no hay laboratorio)

Si queremos aumentar el nivel de confianza...

$$(1-\alpha) = 0.99 \implies z_{0.995} = 2.57$$

entonces

$$\left[\widehat{\overline{x}} \pm \frac{\sqrt{\sigma^2}}{\sqrt{n}} 2.57\right] = [153,81;\ 174,158]$$

menor precisión!!

En economía generalmente se dispone de una única muestra (no hay laboratorio)
Si queremos aumentar el nivel de confianza...

$$(1-\alpha) = 0.99 \implies z_{0.995} = 2.57$$

entonces

$$\left[\widehat{\overline{x}} \pm \frac{\sqrt{\sigma^2}}{\sqrt{n}} 2.57\right] = [153, 81;\ 174, 158]$$

menor precisión!!

Dado $1-\alpha$ podemos aumentar la precisión de dos formas

1. disminuir σ^2 (política de distribución de rentas)
2. aumentar n (tamaño de muestra)

4.2. Distribución de la varianza y la cuasi-varianza muestral

Estimador de la varianza

Sean $X_i \sim \text{iid. N}\left(\mu\,,\,\sigma^2\right)$, con μ y σ^2 desconocidos.

Emplearemos como estimador de $\text{Var}\,(X_i)$ el estadístico:

$$g(X_1,\ldots,X_n) = \sum_{i=1}^{n} \frac{(X_i - \widehat{\overline{x}})^2}{\sigma^2} = \frac{n\widehat{s^2}}{\sigma^2} = \frac{(n-1)\widehat{\mathfrak{s}^2}}{\sigma^2} \sim \chi^2_{(n-1)}$$

donde

$$\widehat{s^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\overline{x}})^2,$$

y

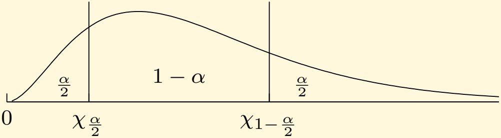
$$\widehat{\mathfrak{s}^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\overline{x}})^2 \quad \text{(insesgado).}$$

Cálculo del intervalo Puesto que

$$\frac{(n-1)\widehat{\mathfrak{s}^2}}{\sigma^2} \sim \chi^2_{(n-1)}$$

Fijado $(1-\alpha)$ buscamos a y b tales que

$$P\left(a \leq \frac{(n-1)\widehat{\mathfrak{s}^2}}{\sigma^2} \leq b\right) = (1-\alpha)$$



$$\text{entonces } \sigma^2 \in \left[\frac{\widehat{\mathfrak{s}^2}(n-1)}{\chi^2_{(n-1,1-\alpha/2)}}, \frac{\widehat{\mathfrak{s}^2}(n-1)}{\chi^2_{(n-1,\alpha/2)}}\right] \text{ con prob. } = (1-\alpha)$$

4.3. Distribución de la media muestral cuando la varianza NO es conocida

ESTIMADOR DE LA ESPERANZA CUANDO LA VARIANZA NO ES CONOCIDA

Sea X, y sean $\mathrm{Var}\,(X)$ y $\mathrm{E}\,(X)$ desconocidas.
Emplearemos como estimador de $\mathrm{E}\,(X)$ el estadístico:

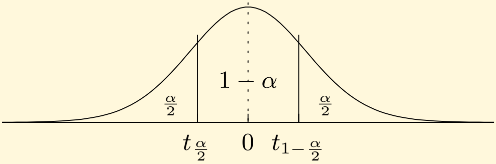
$$g(X_1, \ldots, X_n) = \frac{\widehat{\overline{x}} - \mu}{\sqrt{\widehat{\mathfrak{s}^2}/n}} \sim t_{n-1}\,.$$

Cálculo del intervalo Puesto que

$$\frac{\widehat{\overline{x}}-\mu}{\sqrt{\widehat{\mathfrak{s}^2}/n}} \sim t_{n-1}$$

Fijado $(1-\alpha)$ buscamos a y b tales que

$$P\left(a \leq \frac{\widehat{\overline{x}}-\mu}{\sqrt{\widehat{\mathfrak{s}^2}/n}} \leq b\right) = (1-\alpha)$$



entonces

$$\mu \in \left[\widehat{\overline{x}} \pm\, t_{n-1,\alpha/2} \cdot \sqrt{\widehat{\mathfrak{s}^2}/n}\right] \quad \text{con Prob. } = (1-\alpha).$$

4.4. Otras distribuciones

Sean $X_1, X_2, \ldots, X_n$ con $\mathrm{E}\left(X_i\right) = \mu_i$, con $\mathrm{Var}\left(X_i\right) = \sigma_i^2$ con $f_{X_i}\left(x_i\right)$ cualquiera; entonces si $Y = X_1 + X_2 + \ldots + X_n$

$$\frac{Y - \sum \mu_i}{\sqrt{\sum \sigma_i^2}} \xrightarrow{\mathrm{d}} N(0,1);$$

por lo tanto

$$Y \xrightarrow{\mathrm{d}} N\left(\sum \mu_i,\ \sum \sigma_i^2\right).$$

Tablas resumen de varios estadísticos

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
p	$\frac{\widehat{p}-p}{\sqrt{\frac{\widehat{p}\,\widehat{q}}{n}}}$	$\mathrm{N}\,(0\,,\,1)$	$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}\,\widehat{q}}{n}}$
$p_1 - p_2$	$\frac{(\widehat{p_1}-\widehat{p_2})-(p_1-p_2)}{\sqrt{\frac{\widehat{p_1}\,\widehat{q_1}}{n_1}+\frac{\widehat{p_2}\,\widehat{q_2}}{n_2}}}$	$\mathrm{N}\,(0\,,\,1)$	$\widehat{p_1}-\widehat{p_2} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\,\widehat{q_1}}{n_1}+\frac{\widehat{p_2}\,\widehat{q_2}}{n_2}}$

Cuadro 1: Intervalos de confianza para proporciones; $(q = 1 - p)$

Parámetro	Estadístico pivote	Distribución	Intervalo de confi
μ; σ conocido	$\frac{(\bar{x}-\mu)}{\sqrt{\sigma^2/n}}$	$\mathrm{N}\,(0\,,\,1)$	$\bar{x} \pm z_{\frac{\alpha}{2}} \sqrt{\sigma^2/}$
μ; σ desconocido	$\frac{(\bar{x}-\mu)}{\sqrt{\widehat{\mathfrak{s}^2}/n}}$	t_{n-1}	$\bar{x} \pm t_{\frac{\alpha}{2}} \sqrt{\widehat{\mathfrak{s}^2}/}$
σ^2	$(n-1)\widehat{\mathfrak{s}^2}/\sigma^2$	χ^2_{n-1}	$\frac{\widehat{\mathfrak{s}^2}\,(n-1)}{\chi^2_{1-\frac{\alpha}{2}}},\ \frac{\widehat{\mathfrak{s}^2}\,(n-}{\chi^2_{\frac{\alpha}{2}}}$
$\mu_1-\mu_2$ $\sigma_1,\ \sigma_2$ conocidas	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$	$\mathrm{N}\,(0\,,\,1)$	$\bar{x}_1-\bar{x}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n}+}$
$\mu_1-\mu_2$; $\sigma_1 \neq \sigma_2$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\mathfrak{s}_1^2}{n}+\frac{\mathfrak{s}_2^2}{m}}}$	t_{n+m-2}	$\bar{x}_1-\bar{x}_2 \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\mathfrak{s}_1^2}{n}+}$
$\mu_1-\mu_2$; $\sigma_1=\sigma_2$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n-1)\mathfrak{s}_1^2+(m-1)\mathfrak{s}_2^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)}}$	t_{n+m-2}	$\bar{x}_1-\bar{x}_2 \pm t_{\frac{\alpha}{2}}$ denomina
$\frac{\sigma_2^2}{\sigma_1^2}$	$\frac{\widehat{\mathfrak{s}_1^2}/\sigma_1^2}{\widehat{\mathfrak{s}_2^2}/\sigma_2^2} = \frac{\widehat{\mathfrak{s}_1^2}\ \sigma_2^2}{\widehat{\mathfrak{s}_2^2}\ \sigma_1^2}$	$F_{n-1,m-1}$	$F_a \frac{\widehat{\mathfrak{s}_2^2}}{\widehat{\mathfrak{s}_1^2}} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_b$

5. Trasparencias

Lista de Trasparencias

6. Bibliografía

Novales, A. (1997). *Estadística y Econometría*. McGraw-Hill, Madrid, primera ed. ISBN 84-481-0798-5. 5

Peña, D. (2001). *Fundamentos de Estadística*. Alianza Editorial, Madrid. ISBN 84-206-8696-4. 5

Peña, D. (2002). *Regresión y diseño de experimentos*. Alianza Editorial, Madrid. ISBN 84-206-8695-6. 5

Peña, D. y Romo, J. (1997). *Introducción a la Estadística para la Ciencias Sociales*. McGraw-Hill, Madrid. ISBN 84-481-1617-8. 5

Partes del temario

- Tema 1 IntEctr-T01
- Tema 2 IntEctr-T02
- Tema 3 IntEctr-T03
- Tema 4 IntEctr-T04
- Tema 5 IntEctr-T05
- Tema 6 IntEctr-T06
- Tema 7 IntEctr-T07