# Introducción a los intervalos de confianza.

Mathieu Kessler

Departamento de Matemática Aplicada y Estadística
Universidad Politécnica de Cartagena

Cartagena, Enero 2010

# Guión

# Guión

# Introducción

## El contexto

- Consideramos un fenómeno aleatorio y formulamos un modelo para la variable de interés $X$.
- Pero nos falta ajustar el valor concreto de uno o varios parámetros. *Ejemplo: decidimos modelizar la resistencia de probetas de cemento por una Normal, pero ¿con qué $\mu$ y $\sigma$?*
- Repetimos el experimento varias veces, apuntamos los valores de $X$.
- ¿Cómo usar estos valores para aproximar el (los) parámetro(s)? $\Leftrightarrow$ Estimar los paramétros

## Guión

# Estimación puntual

Consideramos exp. aleatorio, v.a $X$. Hemos decidido un modelo para la distribución de $X$ pero nos falta algún parámetro. Disponemos de una muestra de observaciones de $X$.

## Definición

Cualquier estadístico (es decir, cualquier función de las observaciones de la muestra) diseñado para aproximar el valor de un parámetro $\theta$ del modelo, se llama estimador puntual del parámetro $\theta$.

## Ejemplos de estimadores

| $\theta$ | Estimador |
|---|---|
| $\mu$ | $\bar{X}$, media muestral |
| $\sigma^2$ | $S^2$, varianza muestral |
| $p$ | $\hat{p}$, proporción muestral |

## ¡Un estimador es una variable aleatoria!

- El valor que toma depende de la muestra concreta escogida.

$$\boxed{\text{Es muy útil conocer su distribución muestral}}$$

Puede ser muy difícil deducirla para algunos modelos.

## Propiedades deseables para un estimador

- Estimador insesgado: $\mathbb{E}[\hat{\theta}] = \theta.$
  Ejemplo: vimos $\mathbb{E}[\bar{X}] = \mu$, i.e $\bar{X}$ *es un estimador insesgado de la media* $\mu$.
- Estimador consistente
  Es insesgado y **además**

$$Var(\hat{\theta}) \xrightarrow{n\to\infty} 0.$$

Ejemplo: vimos $Var[\bar{X}] = \frac{\sigma^2}{n}$, i.e $\bar{X}$ *es un estimador consistente de la media* $\mu$.

# Guión

# Estimación por intervalos

## Idea básica

- Para $X \sim \mathcal{N}(\mu, 4)$, supongo que quiero estimar $\mu$.
- Para ello, extraigo una muestra de 4 observaciones y calculo $\bar{X}$.
- Sé que $\bar{X} \sim \mathcal{N}(\mu, 4/4)\bar{X} \sim \mathcal{N}(\mu, 1)$
- $\Rightarrow$ Para el 95% de las muestras que podría extraer, el valor de $\bar{X}$ está entre $\mu - 2$ y $\mu + 2$.
- $\Rightarrow$ Al extraer una muestra, tengo una probabilidad de 0.95 de que $\bar{X}$ esté a menos de 2 unidades de $\mu$.
- Al extraer una muestra, tengo una probabilidad de 0.95 de que $\mu$ esté a menos de 2 unidades de $\bar{X}$.

$$\mu = \bar{X} \pm 2.$$

# Construcción de un IC (Intervalo de Confianza) para $\mu$

## Contexto

V.a. $X \sim \mathcal{N}(\mu, \sigma^2)$. Suponemos que **conocemos** el valor de $\sigma^2$.
Pensamos extraer una muestra de tamaño $n$.

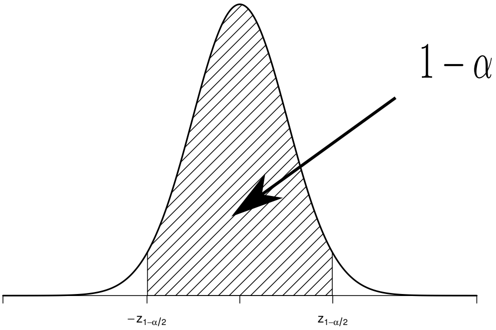
Los pasos que seguimos para la construcción del IC son:

1. Nos fijamos un nivel de riesgo $\alpha$. Nivel de confianza: $1-\alpha$.
   Valores más usados: $\alpha = 0.1$, $0.05$ y $0.01$.
   Niveles de confianza asociados: 90%, 95% y 99% resp.
2. Nos basamos en $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, lo que implica:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1).$$

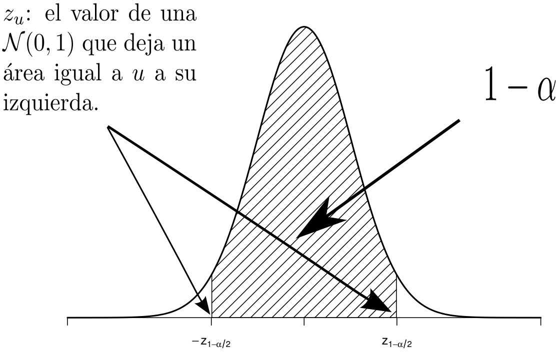
# Construcción de un IC (Intervalo de Confianza) para $\mu$

3. Dibujo en la densidad de $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ una región central que contenga $100(1-\alpha)\%$ del área total



# Construcción de un IC (Intervalo de Confianza) para $\mu$

**3** Dibujo en la densidad de $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ una región central que contenga $100(1-\alpha)\%$ del área total



4 Deduzco

$$\mathbb{P}(-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}) = 1-\alpha,$$

despejo:

$$\mathbb{P}(\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}) = 1-\alpha$$

6 El intervalo de confianza al $100(1-\alpha)\%$ para $\mu$ es

$$\mu \in [\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}; \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}].$$

De manera equivalente:

$$\boxed{\mu = \bar{X} \pm z_{1-\alpha/2}\sigma/\sqrt{n},}$$

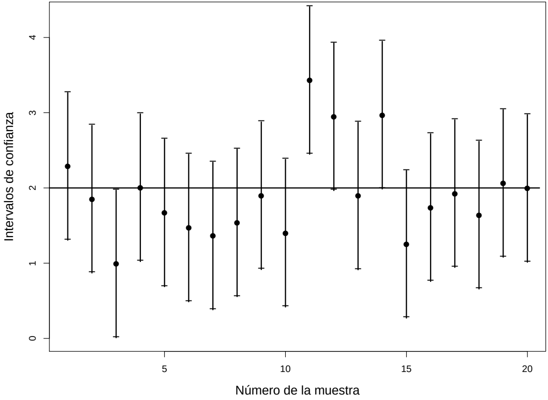
El término $z_{1-\alpha/2}\sigma/\sqrt{n}$ se llama término de error.

## Interpretación

- El intervalo $[\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}; \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}]$ es un intervalo aleatorio.
- Tiene una probabilidad $100(1-\alpha)\%$ de capturar el valor de $\mu$.
- Al extraer una muestra, tengo una probabilidad $1-\alpha$ de que el intervalo que proporcione contenga el valor de $\mu$.
- Al extraer una muestra, corro un riesgo $\alpha$ de que el intervalo que proporcione NO contenga el valor de $\mu$.

# Simulación de 20 muestras de tamaño 4, IC asociados

El auténtico $\mu$ es 2.



# Ejemplo

- Queremos estimar la longitud media de un artículo producido por una máquina.
- Exp: "escoger una pieza al azar en la producción". $X$: "longitud de la pieza escogida".
- Decidimos modelizar la distribución de $X$ por una normal. $X \sim \mathcal{N}(\mu, \sigma^2)$, con $\sigma = 0.05$.
- Extraemos una muestra de 5 artículos:

$$20.1,\ 20.05,\ 20.01,\ 19.95,\ 19.99.$$

- $\bar{x} = 20.02$, $n = 5$. Al 90% de confianza, $\alpha = 0.1$. $z_{0.95} = 1.64$. (tabla o R)
- $\mu = \bar{x} \pm z_{1-\alpha/2}\sigma/\sqrt{n} = 20.02 \pm 1.64\frac{0.05}{\sqrt{5}} = 20.02 \pm 0.04.$

# Comentarios importantes

## ¿Si $X$ no es Normal?

- En los cálculos anteriores hemos usado $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$. Exacto si $X$ es Normal....
- Si $n$ es grande, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ aproximadamente, (Teorema Central del Límite...). $\Rightarrow$ Los IC siguen aprox. válidos.

## Factores que influyen la precisión del intervalo

Margen de error: $z_{1-\alpha/2}\sigma/\sqrt{n}$.

1. $n \uparrow, \Rightarrow$ precisión $\uparrow$
2. $\sigma \uparrow \Rightarrow$ precisión $\downarrow$
3. confianza $\uparrow \Rightarrow$ precisión $\downarrow$

# Determinación del tamaño muestral

## Contexto

Antes de extraer la muestra:

- Tenemos decidido el valor de $\sigma$.
- Tenemos decidido la confianza con la que trabajamos.
- Tenemos decidido el margen de error máximo *max* que estamos dispuesto a cometer

¿Qué tamaño de la muestra debemos escoger?

$$\text{Margen de error: } z_{1-\alpha/2}\sigma/\sqrt{n} \qquad \leq max.$$

Despejamos $n$...

# Ejemplo

*La medición de la conductividad de un material sigue una distribución que modelizamos por una Normal con desviación típica $\sigma = 0.5$. Quiero construir un intervalo de confianza al 95% para el valor promedio proporcionado de la conductividad pero quiero que el error cometido sea menor de 0.3. ¿cuántas veces deberé repetir la medición?*

Busco $n$ tal que $z_{1-\alpha/2}\sigma/\sqrt{n} \leq 0.3$, sabiendo que $\sigma = 0.5$, y $\alpha = 0.05$.

Obtengo

$$1.96\frac{0.5}{\sqrt{n}} \leq 0.3,$$

$$n \geq \left(\frac{1.96 \cdot 0.5}{0.3}\right)^2 \simeq 10.67.$$

Habrá por lo tanto que realizar 11 mediciones.

# ¿Cómo construir un IC si desconocemos $\sigma$?

## Nota importante

- *La construcción anterior supone que como parte de nuestro modelado, ANTES de extraer una muestra, hemos fijado un valor de $\sigma$ en $X \sim \mathcal{N}(\mu, \sigma^2)$.*
- Este supuesto no es realista... Es más normal que sólo especifiquemos en nuestro modelado: $X \sim \mathcal{N}(\mu, \sigma^2)$ y que tanto $\mu$ como $\sigma$ sean desconocidos.
- En este caso debemos estimar tanto $\mu$ (con la media muestral) como $\sigma$ (con la desviación típica muestral)...

# ¿Cómo construir un IC si desconocemos $\sigma$?

## Resultado fundamental

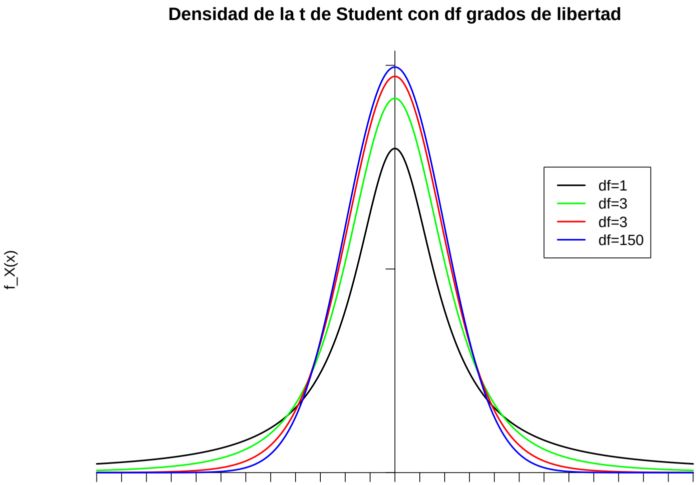
$X \sim \mathcal{N}(\mu, \sigma^2)$, con $\mu$ y $\sigma$ desconocidos....

Al extraer una muestra de tamaño $n$, consideramos el estadístico muestral:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

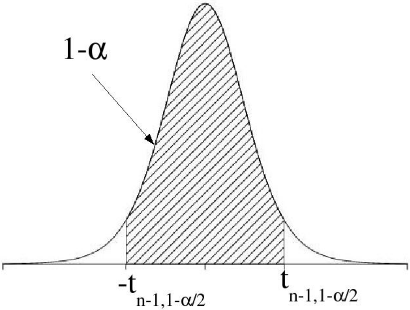
Sigue una distribución t de Student con $n - 1$ grados de libertad.

# Densidad de la t de Student



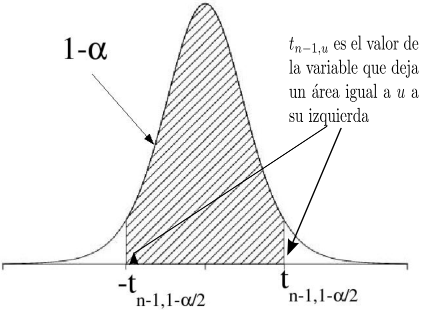
# Construcción del IC para $\mu$ si $\sigma$ es desconocida

1. Nos fijamos un nivel de riesgo $\alpha$. Nivel de confianza: $1-\alpha$.
2. Nos basamos en
$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$
3. Dibujo una región central que represente el $100(1-\alpha)\%$ del área total



# Construcción del IC para $\mu$ si $\sigma$ es desconocida

1. Nos fijamos un nivel de riesgo $\alpha$. Nivel de confianza: $1-\alpha$.
2. Nos basamos en
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$
3. Dibujo una región central que represente el $100(1-\alpha)\%$ del área total



# Construcción del IC para $\mu$ si $\sigma$ es desconocida

- Deducimos

$$\mathbb{P}(-t_{n-1,1-\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{n-1,1-\alpha/2}) = 1 - \alpha.$$

- Despejamos $\mu$: el intervalo de confianza al $100(1-\alpha)\%$ para $\mu$ es

$$\mu \in [\bar{X} - t_{n-1,1-\alpha/2} S/\sqrt{n}; \bar{X} + t_{n-1,1-\alpha/2} S/\sqrt{n}].$$

- Se escribe también

$$\boxed{\mu = \bar{X} \pm t_{n-1,1-\alpha/2} S/\sqrt{n}},$$

- El término $t_{n-1,1-\alpha/2} S/\sqrt{n}$ es el término de error.

# Ejemplo

*A la hora de realizar mediciones de una cantidad, introduzco la variable $X$:"Valor proporcionado por el aparato", y decido modelizar la distribución de los valores que puede tomar por una Normal, media $\mu$ y varianza $\sigma^2$, ambas desconocidas.*

- Decido construir un IC para $\mu$ al 95% de confianza, usando tres mediciones.
- Realizo las tres mediciones y obtengo: $10,23$, $10,24$ y $10,26$.
- $\mu = \bar{x} \pm t_{n-1,1-\alpha/2} S/\sqrt{n}$,
  necesitamos $\bar{x} = 10.2433$ y $S^2 = 0.0002333$. Además $t_{2,0.975} = 4.301$ (R)
- $\mu = 10.2433 \pm 4.303\sqrt{0.000233}/\sqrt{3} = 10.2433 \pm 0.03793$.