Motivación

X espesor de una capa asfáltica en centímetros. Se dispone de una muestra de tamaño 9 de X.

$$\begin{array}{ccccc} 3{,}97 & 4{,}78 & 4{,}87 & 4{,}92 & 5{,}97 \\ 4{,}27 & 6{,}03 & 6{,}61 & 3{,}81 & \end{array}$$

Sabemos que X es normal con varianza 1: $X \in N(m, 1)$, y queremos hacer una estimación de la media m.

(La hipótesis de varianza conocida no es realista.)

Por estimación puntual sería razonable tomar $\hat{m} = \overline{X} \approx 5{,}026$

"Desenfocar" la estimación puntual dando un intervalo en el que se pueda controlar la incertidumbre de la estimación que estamos haciendo: Intervalo de confianza.

Motivación

Tenemos una muestra de tamaño 9 de una variable $X \in N(m, 1)$. Sabemos que bajo esas hipótesis

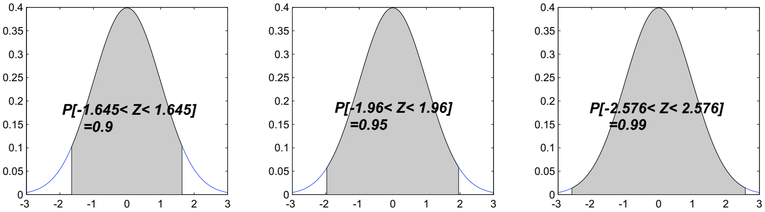
$$\overline{X} \in N(m, \frac{1}{9})$$

y estandarizando

$$U = \frac{\overline{X} - m}{\sqrt{1/9}} \in N(0, 1)$$

Motivación

El siguiente paso es fijar intervalos centrados en 0 y con diferentes probabilidades con respecto a la normal estándar.



Por ejemplo si nos fijamos en el segundo intervalo:

$$P\left[-1{,}96 \leq \frac{\overline{X}-m}{\sqrt{1/9}} \leq 1{,}96\right] = 0{,}95$$

que se puede escribir como

$$P[m - 0{,}6533 \leq \overline{X} \leq m + 0{,}6533] = 0,95$$

pero a nosotros nos interesa escrito de esta forma:

$$P[\overline{X} - 0{,}6533 \leq m \leq \overline{X} + 0{,}6533] = 0,95$$

Motivación

$P[\overline{X} - 0{,}6533 \leq m \leq \overline{X} + 0{,}6533] = 0,95$

Sustituimos el valor de $\overline{X} = 5{,}026$ obtenido de la muestra y nos queda

$\cancel{P[4{,}3722 \leq m \leq 5{,}6788] = 0,95}$

Se dice más bien que $[4{,}3722,\ 5{,}6788]$ es un **intervalo de confianza** $95\,\%$ **sobre la media** m.

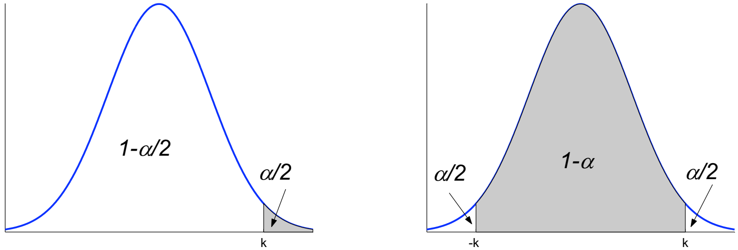
Intervalos de confianza sobre la media en poblaciones normales de varianza conocida: Intervalos por los dos lados

$X \in N(m, \sigma^2)$ con σ^2 conocida. Queremos calcular un intervalo de confianza $1-\alpha$ sobre m a partir de una muestra $(x_1, \cdots, x_n)$ de tamaño n de X.

$1-\alpha$: confianza. α : significación. En el ejemplo anterior, $\alpha = 0{,}05$.

Valores habituales para α : 0,01 0,05 0,1.

Buscamos a tal que $P[-a \leq U \leq a] = 1-\alpha$ donde U es una normal estándar. Esto quiere decir que $F_U(a) = 1-\alpha/2$



Intervalos de confianza sobre la media en poblaciones normales de varianza conocida: Intervalos por los dos lados

$P[-a \leq U \leq a] = 1 - \alpha$ donde U es una normal estándar. Si sustituimos

$P[-a \leq \frac{\overline{X}-m}{\sigma/\sqrt{n}} \leq a] = 1 - \alpha$ llegamos a que

$P[\overline{X} - a\frac{\sigma}{\sqrt{n}} \leq m \leq \overline{X} + a\frac{\sigma}{\sqrt{n}}] = 1 - \alpha$ o dicho de otra forma,

- $[\overline{X} - a\frac{\sigma}{\sqrt{n}}, \quad \overline{X} + a\frac{\sigma}{\sqrt{n}}]$ **es un intervalo de confianza** $1 - \alpha$ **sobre** m **(donde** $F_U(a) = 1 - \alpha/2$**)**

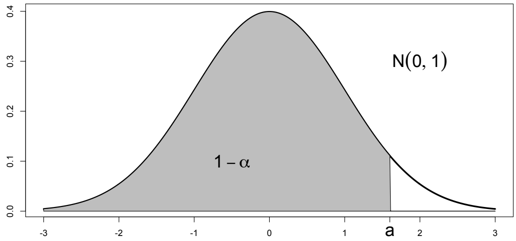
Con la muestra de antes ($\overline{X} = 5{,}025556;\ n = 9;\ \sigma^2 = 1$):

$1-\alpha$	α	$1-\alpha/2$	a	Extr. inf.	Extr. sup.
0.9	0.1	0.95	1.644854	4.477271	5.573840
0.95	0.05	0.975	1.959964	4.372234	5.678877
0.99	0.01	0.995	2.575829	4.166946	5.884165

Intervalos de confianza sobre la media en poblaciones normales de varianza conocida: Intervalos por un solo lado

$X \in N(m, \sigma^2)$ con σ^2 conocida. Queremos calcular un intervalo de confianza $1-\alpha$ sobre m **por la izquierda** a partir de una muestra $(x_1, \cdots, x_n)$ de tamaño n de X, es decir, llegar a una expresión del tipo $\quad P[??? \leq m] = 1-\alpha$

Buscamos a tal que $P[U \leq a] = 1-\alpha$ donde U es una normal estándar. Esto quiere decir que $F_U(a) = 1-\alpha$ (y no $1-\alpha/2$ como en el caso bilateral)



Intervalos de confianza sobre la media en poblaciones normales de varianza conocida: Intervalos por un solo lado

$P[U \leq a] = 1 - \alpha$ donde U es una normal estándar. Si sustituimos

$P[\frac{\overline{X}-m}{\sigma/\sqrt{n}} \leq a] = 1 - \alpha$ llegamos a que

$P[\overline{X} - a\frac{\sigma}{\sqrt{n}} \leq m] = 1 - \alpha$ o dicho de otra forma,

- $[\overline{X} - a\frac{\sigma}{\sqrt{n}}, \quad +\infty)$ **es un intervalo de confianza** $1 - \alpha$ **por la izquierda sobre** m **(donde** $F_U(a) = 1 - \alpha$**)**
- Análogamente $(-\infty, \quad \overline{X} + a\frac{\sigma}{\sqrt{n}}]$ **es un intervalo de confianza** $1 - \alpha$ **por la derecha sobre** m **(donde** $F_U(a) = 1 - \alpha$**)**

Intervalos de confianza sobre la media en poblaciones normales de varianza conocida: Intervalos por un solo lado

Intervalos por la izquierda $[\overline{X} - a\frac{\sigma}{\sqrt{n}}, \quad +\infty)$, con la muestra de antes ($\overline{X} = 5{,}025556;\ n = 9;\ \sigma^2 = 1$):

$1-\alpha$	α	a	Extr. inf.	Extr. sup.
0.9	0.1	1.281552	4.598372	$+\infty$
0.95	0.05	1.644854	4.477271	$+\infty$
0.99	0.01	2.326348	4.250106	$+\infty$

Intervalos por la derecha $(-\infty, \quad \overline{X} + a\frac{\sigma}{\sqrt{n}}]$, con la muestra de antes ($\overline{X} = 5{,}025556;\ n = 9;\ \sigma^2 = 1$):

$1-\alpha$	α	a	Extr. inf.	Extr. sup.
0.9	0.1	1.281552	$-\infty$	5.452739
0.95	0.05	1.644854	$-\infty$	5.573840
0.99	0.01	2.326348	$-\infty$	5.801005

Aplicación de las fórmulas

Para aplicar estas fórmulas no es estrictamente necesario que la variable X sea normal sino que lo sea $\overline{X}$. Esto sucede

- exactamente si X es normal (propiedad regenerativa), independientemente del tamaño muestral. Éste es el caso que hemos supuesto.
- aproximadamente si la muestra es grande (Teorema del Límite Central), aunque X no sea normal.

La hipótesis de varianza conocida no se da en la práctica. Veremos más tarde que para muestras grandes se pueden aplicar estas fórmulas con carácter aproximado tomando $\sigma_X^2 \approx S_X^2$.