

## Rincón de la Bioestadística

# Docimasia de hipótesis

Gabriel Cavada Ch.<sup>1</sup>

<sup>1</sup>División de Bioestadística, Escuela de Salud Pública, Universidad de Chile.

## Hypothesis testing

Hipótesis estadística es una afirmación respecto de una característica poblacional (la forma de ella o el valor de sus parámetros); esta sentencia puede ser “docimada” (probada) usando una muestra aleatoria extraída de esa población.

En muchas ocasiones es necesario decidir entre una afirmación de la forma  $\theta = \theta_0$  (Hipótesis nula) u otra que puede tomar las siguientes formas  $\theta \neq \theta_0, \theta > \theta_0, \theta < \theta_0$  (Hipótesis alternativa). En símbolos:

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

ó

$$H_1 : \theta > \theta_0$$

ó

$$H_1 : \theta < \theta_0$$

Para desarrollar un procedimiento que permita decidir acerca de  $H_0$ , dicha decisión será tomada en base a información muestral, la que está sujeta a errores probables debido a que no se sabe con certeza cómo es la naturaleza, y sólo disponemos de una percepción de ella. Cruzando este efecto con la decisión tenemos:

		Estado de la naturaleza	
		$H_0$ es Verdad	$H_0$ es Falsa
Percepción de la naturaleza	Rechazar $H_0$	Error tipo I	Decisión correcta
	No rechazar $H_0$	Decisión correcta	Error tipo II

Desearíamos que los errores no se cometieran, pero como la decisión será tomada bajo incertidumbre, sólo podemos aspirar a que la probabilidad de cometerlos sea pequeña.

La filosofía para “docimar” consiste en suponer que  $H_0$  es verdadera, hasta encontrar evidencia muestral suficiente que permita decir lo contrario; si esta evidencia no existe no podemos dudar de la afirmación contenida en  $H_0$ . Así el error más grave que se puede cometer es el Error tipo I, que es el que tratamos de controlar. Llamando:

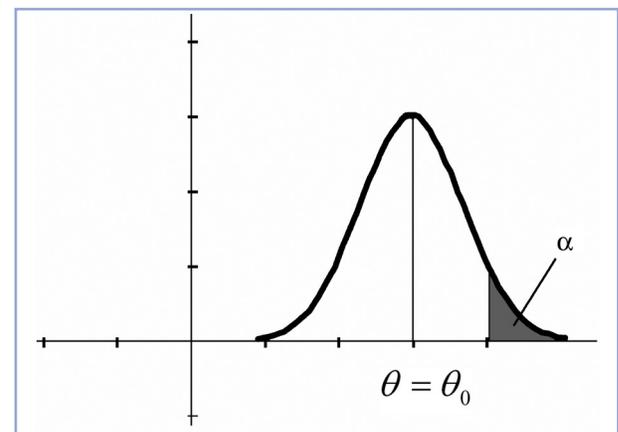
$\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ es Verdad})$ , tamaño del Error tipo I

$\beta = P(\text{No rechazar } H_0 \mid H_0 \text{ es Falsa})$ , tamaño del Error tipo II

Nos interesa que  $\alpha$  sea pequeño;  $\alpha$  se llama significación de la “dócima” y  $1-\beta$  se llama potencia de la “dócima”, potencia que depende de la hipótesis alternativa que estemos proponiendo y puede interpretarse como el grado de credibilidad que asignamos a la hipótesis alternativa.

Se llama estadística de prueba E, a una función que contenga el parámetro de interés (que se desea docimar) y toda la información muestral. Deseablemente la estadística de prueba, bajo la hipótesis nula, debe seguir una distribución conocida de probabilidades.

Se llama región crítica o de rechazo a aquella porción de los reales para la cual la probabilidad de que E esté en ella, considerando la veracidad de  $H_0$ , sea menor que  $\alpha$



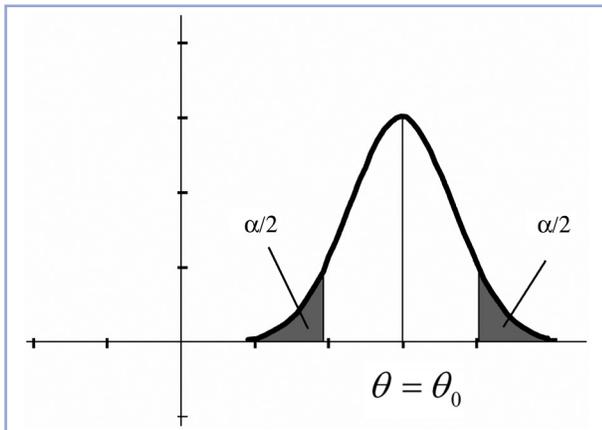
• Una “dócima” de la forma siguiente:

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

## Rincón de la Bioestadística

se llama de “dos colas”, pues la región de rechazo se compone de dos porciones inconexas de los reales, que se muestran en el siguiente gráfico:



- Una “dócima” de la siguiente forma:

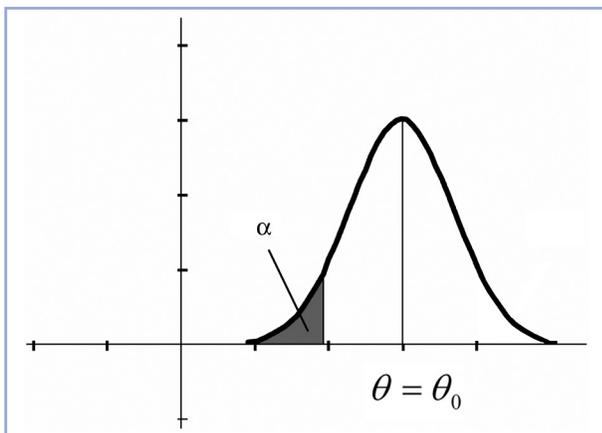
$$H_0 : \theta = \theta_0$$

$$H_1 : \theta > \theta_0$$

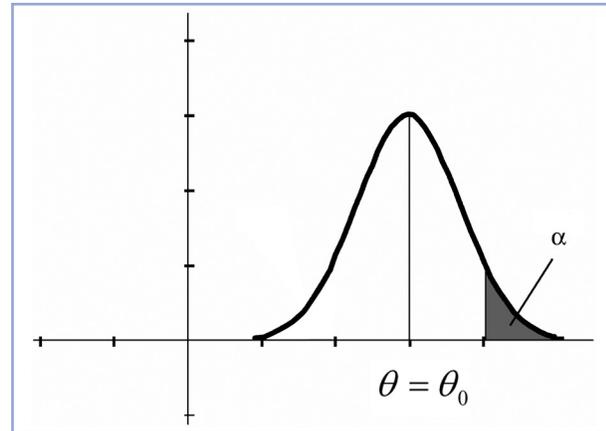
ó

$$H_1 : \theta < \theta_0$$

se llama de “una cola” pues la región de rechazo se compone de una porción conexas de los reales, como se muestra a continuación:



$$H_1 : \theta < \theta_0$$



$$H_1 : \theta > \theta_0$$

Desde la entronización de la computación, se ha extendido más el uso del valor “p”, que el tradicional nivel de significación. El valor de “p” no es más que el valor exacto del tamaño del Error I. Sin embargo, para su correcta interpretación debemos recordar el supuesto básico de la docimacia de hipótesis que es suponer como verdadera la hipótesis nula. Así, el valor “p” es la probabilidad de haber observado la evidencia que tenemos, en el supuesto que la hipótesis nula fuese cierta. Por ejemplo, si en un ensayo clínico se compara un tratamiento activo contra un placebo, las hipótesis estadísticas planteadas serán:

$H_0$ : El Efecto del Tratamiento “es igual” al Efecto del Placebo.

$H_1$ : El Efecto del Tratamiento “es mejor” que el Efecto del Placebo.

Y si al realizar el experimento se rechaza  $H_0$  a favor de  $H_1$ , con un valor  $p = 0,001$ , la interpretación de este valor es que, si verdaderamente “El Efecto del Tratamiento es igual al Efecto del Placebo”, la probabilidad de haber observado el resultado del ensayo es de un 0,1%. Como no se conoce el verdadero estado de la naturaleza y nuestra percepción de ella está basada en nuestra evidencia, tendemos a creer que  $H_0$  no es cierta y que lo verdadero es lo sostenido en  $H_1$ .

Es decir, una “dócima” de hipótesis no demuestra una hipótesis de investigación, sólo es capaz de asignar grados de credibilidad a nuestras hipótesis de investigación. Es decir, es demasiado pretensioso y abusivo afirmar “está demostrado que el tratamiento activo es mejor que el placebo”; lo correcto es afirmar que “se tiene suficiente evidencia para postular que el tratamiento activo es mejor que el placebo”.

La demografía afirma que la probabilidad de concebir una mujer es aproximadamente de un 50%; si una pareja ha concebido 10 hijos y todos ellos fueron varones, la probabilidad de que esto ocurriera según la demografía es:  $p = 0,5^{10} = 0,00097656$ . ¿Cómo convencer a esta una pareja que la demografía está en lo cierto?