TEMA 8. Contraste de hipótesis de la media

Alicia Nieto Reyes

BIOESTADÍSTICA

Introducción al Contraste de Hipótesis

Vamos a dar respuesta a preguntas del tipo:

- ¿Es realmente cierto que la incidencia media de la sarcoidosis es baja en España?
- Tenemos un tratamiento beneficioso para la hipertensión severa, ¿es también beneficioso para la hipertensión moderada? [Smith et al. 1979]

En un problema de contraste de hipótesis:

- Existe una teoría preconcebida relativa a una característica de la población
- Entonces hay 2 teorías implicitas:
  1. La que propone el experimentador
  2. La negación de esta hipótesis
- Las denotamos por
  1. H_a, la hipótesis alternativa
  2. H_0, la hipótesis nula

Objetivo: Poder rechazar H_0 y, como consecuencia, aceptar H_a

Ejemplo: Tenemos cierto tratamiento que es beneficioso para la hipertensión severa; pero que en principio no lo es para la hipertensión moderada

- Una persona afirma que si es beneficioso para la hipertensión moderada
- Como no queremos lanzar al mercado un producto a menos que estemos seguros de sus efectos positivos

  H_0 : el tratamiento no tiene ningun efecto positivo en pacientes con hipertensión moderada

  H_a : el tratamiento contribuye a disminuir la tensión en pacientes con hipertensión moderada

- Población: pacientes con hipertensión moderada
  Variable: *tensión después de 12 meses de tratamiento menos tensión al inicio del tratamiento.* Suponemos que sigue una distribución normal de media μ y varianza σ^2, desconocidas.
- Entonces: $H_0 : \mu \geq 0$ $H_a : \mu < 0$

Error de tipo I y II

Las decisiones posibles que se pueden tomar una vez seleccionada una muestra son:

		Estado real	
		H_0 cierta	H_a cierta
Decisión	Rechazar H_0	Error de tipo I	Decisión correcta
tomada	Dejar de rechazar H_0	Decisión correcta	Error de tipo II

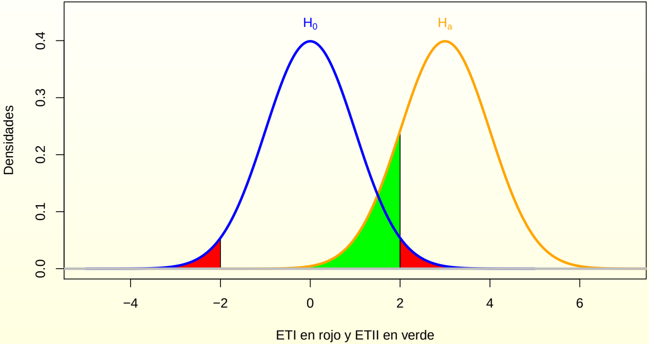
Si pretendemos anular uno de los dos errores, pasamos a no poder controlar el otro

Solución: Encontrar un equilibrio

Ejemplo: Queremos que el error de tipo I sea bajo porque no queremos concluir que el tratamiento funciona cuando no lo hace

Fijamos que la probabilidad de que ocurra un error de tipo I sea $\leq 0{,}05$

Error de tipo I y II



En un determinado estudio H_0 y H_a son como en el dibujo.
Tomamos una muestra y operando obtenemos un valor z:

- si $z \in (-\infty, -2] \cup [2, +\infty)$, rechazamos H_0
- si $z \in (-2, 2)$, dejamos de rechazar H_0

Nota: Se toma el valor 2 con probabilidad cero

- cometemos ETI si H_0 es cierta y $z \in (-\infty, -2] \cup [2, +\infty)$
- cometemos ETII si H_a es cierta y $z \in (-2, 2)$

Contraste de Hipótesis de la Media de una Variable Normal

Antes de llevar a cabo el experimento uno tiene una idea preconcebida sobre el valor de la media

Denotamos por μ_0 a este valor preconcebido

Nos podemos encontrar ante tres tipos de casos:

$H_0 : \mu \geq \mu_0$	$H_0 : \mu \leq \mu_0$	$H_0 : \mu = \mu_0$
$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$	$H_a : \mu \neq \mu_0$

Ejemplo 1: Tenemos cierto tratamiento que es beneficioso para la hipertensión severa y queremos ver si lo es para la hipertensión moderada

$$H_0 : \mu \geq 0 \qquad H_a : \mu < 0$$

Ejemplo 2: La incidencia media de la sarcoidosis es de 19 casos por cada 100.000 habitantes en mujeres. Se cree que en Suecia esta incidencia es mayor. Para demostralo hay que contrastar:

$$H_0 : \mu = 0,19 \qquad H_a : \mu > 0,19$$

Ejemplo 3: El promedio de proteínas en sangre de un adulto sano es de 7,25g/dl. En un análisis de sangre el técnico esta contrastando

$$H_0 : \mu = 7,25 \qquad H_a : \mu \neq 7,25$$

Contraste de Hipótesis de la Media de una Variable Normal

Los 3 ejemplos pueden realizarse para varianza conocida o desconocida
Supongamos que en los siguientes ejemplos es desconocida

Ejemplo 1: Tenemos cierto tratamiento que es beneficioso para la hipertensión severa y queremos ver si lo es para la hipertensión moderada

$$H_0 : \mu \geq 0 \qquad\qquad H_a : \mu < 0$$

Para realizar este test utlizamos el hecho de que bajo H_0,

el estadístico
$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}$$

Tomamos una muestra con $n = 171$; obtenemos $\bar{X} = -9,85$ y $S = 11,52$

Entonces,
$$\frac{\sqrt{171}(-9,85 - 0)}{11,52} = -11.18103$$

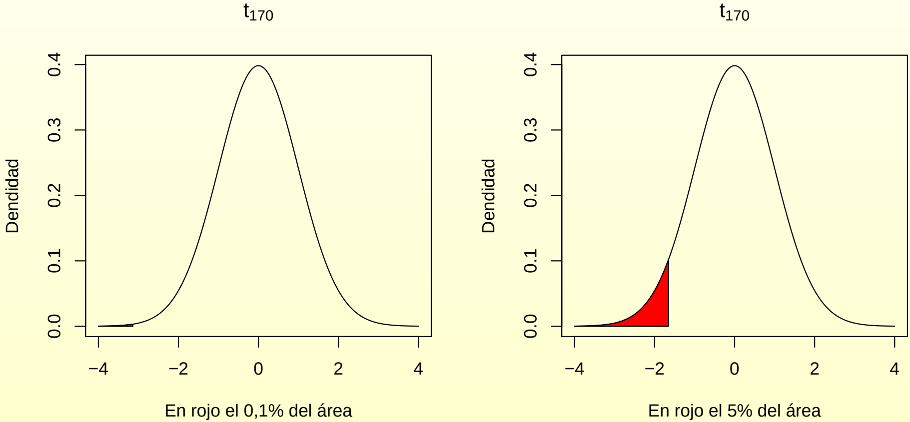
¿Es este valor inusualmente pequeño?

Se denomina $p-$valor a la probabilidad de observar un valor tan extremo o más que aquel realmente obtenido:

$$p = P[t_{170} \leq -11.18103] = 2.086769 * 10^{-22} < 0,001$$

$$p = P[t_{170} \leq -11.18103] = 2.086769 * 10^{-22} < 0,001$$

Como $P[t_{170} \leq -3.138863] = 0,001$,
la Región de Rechazo es $(-\infty, -3.138863]$



Contraste de Hipótesis de la Media de una Variable Normal

- Rechazamos la hipótesis nula al nivel de significación $\alpha = 0,001$

  Recuerda: el nivel de confianza es 1-α

  Para rechazar H_0 bastaría que nos hubiesemos fijado en uno de los dos siguientes hechos:

  - $p < 0,001$
  - $2.086769 * 10^{-22} \in (-\infty, -3.138863]$

- I.e., Afirmamos que el tratamiento tiene efectos beneficiosos para la hipertensión leve al nivel de significación $\alpha = 0,001$

- No decimos que H_0 sea falsa, pero si es cierta, hemos obtenido una muestra que solo se obtiene 1 de cada 1000 veces

Si $p \geq 0,05$ no se rechaza H_0

Ejemplo 2: La incidencia media de la sarcoidosis es de 19 casos por cada 100.000 habitantes en mujeres. Se cree que en Suecia esta incidencia es mayor. Para demostralo hay que contrastar:

$$H_0 : \mu = 0,19 \qquad\qquad H_a : \mu > 0,19$$

Tomada una muestra de tamaño 15, supongamos que obtenemos $\bar{X} = 60$ $S = 1,1$

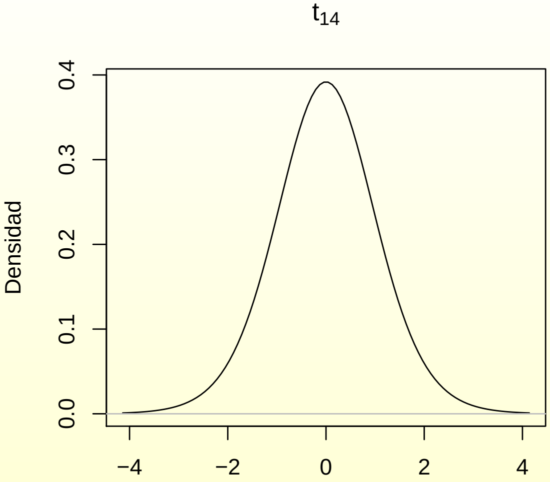
El valor observado del estadístico es

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} = \frac{\sqrt{15}(60 - 19)}{1,1} = 144,357$$

Nota que $\dfrac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{14}$

Entonces: $p = P[t_{14} \geq 144,357] \approx 0$

Contraste de Hipótesis de la Media de una Variable Normal



Como $p = P[t_{14} \geq 144,357] \approx 0$
Rechazamos H_0 (a nivel 0,001)

Ejemplo 3: El promedio de proteínas en sangre de un adulto sano es de 7,25g/dl. En un análisis de sangre el técnico esta contrastando

$$H_0 : \mu = 7,25 \qquad H_a : \mu \neq 7,25$$

Se realizan 10 análisis de sangre sobre un determinado paciente a lo largo de varios días

$$7,215 \quad 7,276 \quad 7,241 \quad 7,265 \quad 7,308 \quad 7,245 \quad 7,283 \quad 7,307 \quad 7,273 \quad 7,219$$

Obtenemos: $\bar{X} = 7,263$ $\qquad$ $S = 0,033$

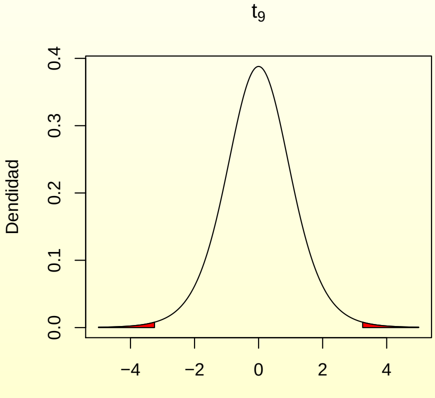
El valor observado del estadístico es

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} = \frac{\sqrt{10}(7,263 - 7,25)}{0,033} = 1,246$$

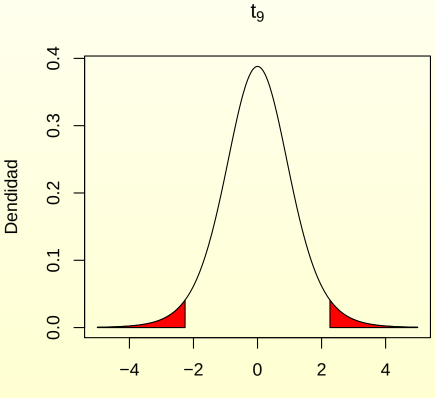
Contraste de Hipótesis de la Media de una Variable Normal

Recuerda

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} = 1,246$$



En rojo el 1% del área



En rojo el 5% del área

Contraste de Hipótesis de la Media de una Variable Normal

Tenemos que

- $P[-4,781 \leq t_9 \leq 4,781] = 1 - 0,001$
  Región de rechazo $(-\infty, -4,781] \cup [4,781, \infty)$
  No tenemos evidencia suficiente para rechazar H_0 a nivel $\alpha = 0,001$
- $P[-3,250 \leq t_9 \leq 3,250] = 1 - 0,01$;
  Región de rechazo $(-\infty, -3,250] \cup [3,250, \infty)$
  No tenemos evidencia suficiente para rechazar H_0 a nivel $\alpha = 0,01$
- $P[-2,262 \leq t_9 \leq 2,262] = 1 - 0,05$;
  Región de rechazo $(-\infty, -2,262] \cup [2,262, \infty)$
  No tenemos evidencia suficiente para rechazar H_0 a nivel $\alpha = 0,05$

No tenemos evidencia suficiente para rechazar H_0