ESTADISTICA.AR

Pruebas de hipótesis: la lógica

Estadística · Semana 10 · Unidad 6: Pruebas de hipótesis

PREGUNTA PARA LA CLASE

Una señora dice que distingue si en la taza se sirvió primero la leche o primero el té. Se le dan 8 tazas, 4 de cada tipo, en orden al azar, y acierta todas. ¿Le creen?

Hoy

1

La lógica de una prueba

¿Qué tan raro sería esto si fuera casualidad?

2

Hipótesis y errores

Las dos formas de equivocarse

3

El p-valor y la región crítica

Dos formas de decidir lo mismo

4

Pruebas para una proporción y una media

Las dos pruebas de hoy

5

Leer bien un resultado

Significativo no quiere decir importante

1

La lógica de una prueba

¿Qué tan raro sería esto si fuera casualidad?

La primera prueba de hipótesis (1710)

John Arbuthnot contó los bautismos de Londres de 1629 a 1710: en los 82 años hubo más varones que mujeres.

82

años seguidos con más varones

1/2

la probabilidad de un año así, si los dos sexos fueran igual de probables

(1/2)^82

la de 82 años seguidos: un número con 24 ceros después de la coma

Fuente: Arbuthnot (1710), Philosophical Transactions of the Royal Society 27: 186-190.

La dama que prueba el té

Si la señora adivina, cuántas de las 4 tazas con la leche primero identifica bien.

Aciertos	Maneras de elegir	Probabilidad si adivina
0	1	1/70 = 0,014
1	16	16/70 = 0,229
2	36	36/70 = 0,514
3	16	16/70 = 0,229
4	1	1/70 = 0,014
Total	70	1

Fuente: Fisher (1935), The Design of Experiments.

La lógica, en tres pasos

Suponer que es casualidad

Que la señora adivina, que los sexos son igual de probables: la hipótesis nula.

Calcular qué tan raro es

La probabilidad de un resultado como el observado, o más extremo, si fuera casualidad.

Decidir

Si es muy improbable, se rechaza la casualidad. Si no, no hay evidencia contra ella.

2

Hipótesis y errores

Las dos formas de equivocarse

Las dos hipótesis

Hipótesis nula, H_0

La de «no hay efecto»: igualdad, ausencia de diferencia, casualidad.

Es la que se pone a prueba.

Hipótesis alternativa, H_1

Lo que se quiere mostrar: que hay un efecto, una diferencia.

Se acepta solo si los datos la apoyan con fuerza.

Las dos hablan del parámetro de la población (μ, p), nunca de la muestra: la muestra ya se conoce.

Dos formas de equivocarse

	Se rechaza H_0	No se rechaza H_0
H_0 es cierta	Error de tipo I (probabilidad α)	Decisión correcta
H_0 es falsa	Decisión correcta (potencia, 1 − β)	Error de tipo II (probabilidad β)

Nunca se sabe en qué fila se está: si se supiera, no haría falta la prueba.

α, β y la potencia

Nivel de significación, α

La probabilidad de error de tipo I que se acepta. Se fija **antes**: 0,05 es lo habitual.

Con la alarma: qué tan seguido se tolera que suene sin fuego.

β

La probabilidad de error de tipo II. Depende de cuánto se aleje la realidad de H_0 y del tamaño de muestra.

No se fija: se calcula para un valor concreto de H_1.

La potencia, 1 − β

La probabilidad de detectar un efecto que existe.

Crece con n y con el tamaño del efecto.

3

El p-valor y la región crítica

Dos formas de decidir lo mismo

La moneda de Kerrich

John Kerrich tiró una moneda 10.000 veces mientras estaba preso en un campo de internación durante la Segunda Guerra Mundial.

5.067

caras en 10.000 tiradas: $\hat{p} = 0{,}5067$

0,005

el error estándar si la moneda es equilibrada: $\sqrt{(0{,}5 \cdot 0{,}5 / 10.000)}$

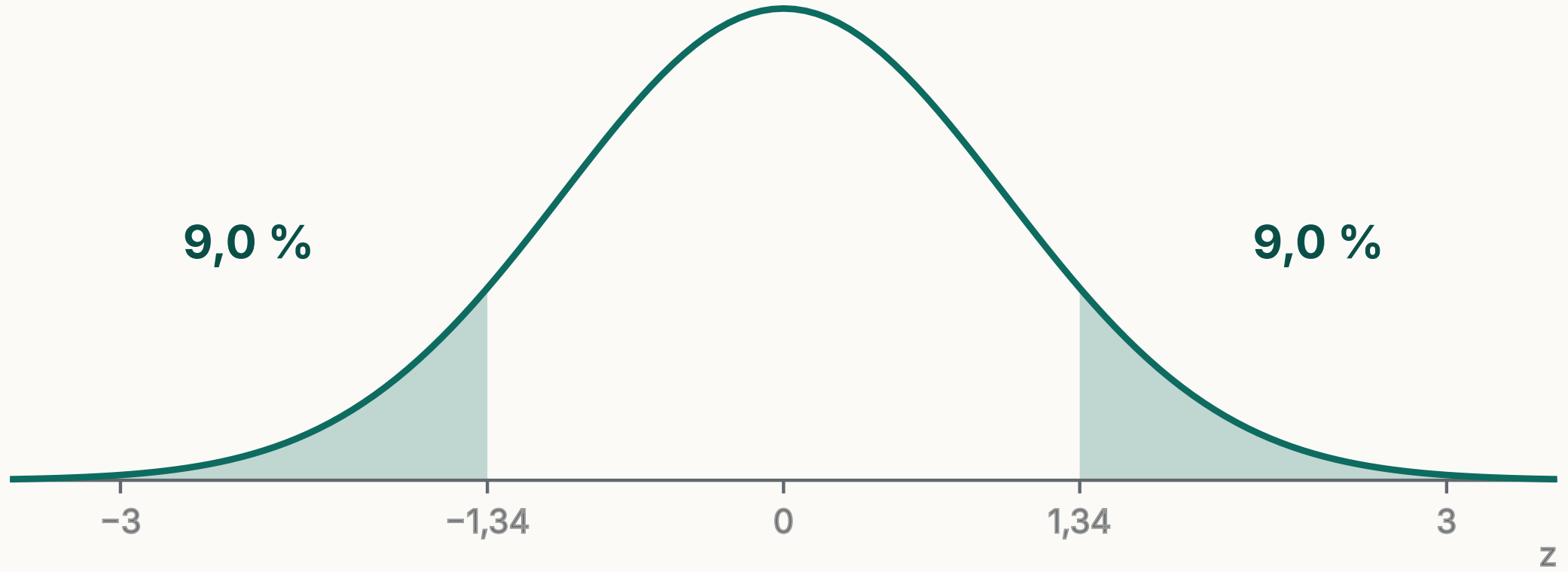
$z = 1{,}34$

$(0{,}5067 - 0{,}5) / 0{,}005$: a cuántos errores estándar de 0,5 quedó

Fuente: Kerrich (1946), An Experimental Introduction to the Theory of Probability.

El p-valor

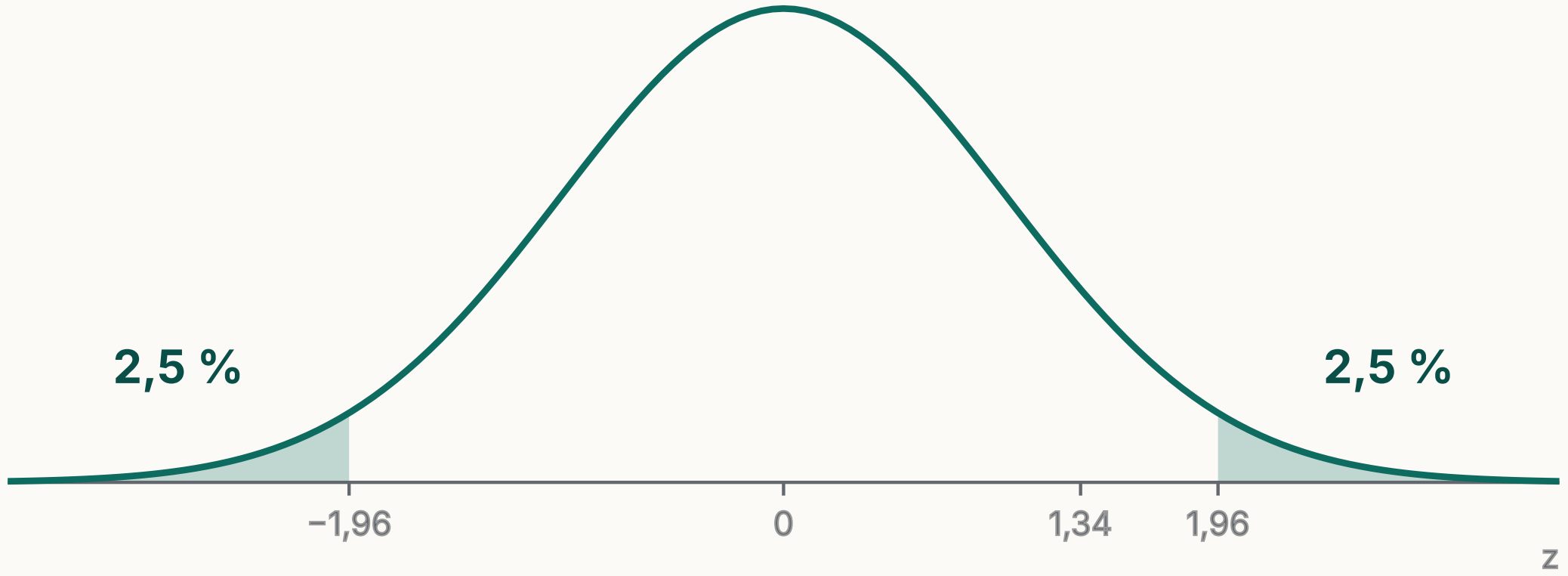
Si la moneda es equilibrada, ¿qué probabilidad hay de quedar tan lejos de 0,5 como Kerrich, o más? $P(|Z| \geq 1{,}34) = 2 \cdot 0{,}0901 = 0{,}1802$



Las dos colas, porque H_1 es $p \neq 0{,}5$: un desvío hacia las cecas sería tan sospechoso como hacia las caras.

La región crítica

Con α = 0,05 y una prueba bilateral, se rechaza H_0 si z cae más allá de ±1,96. El z de Kerrich, 1,34, queda adentro.



Es la otra forma de decidir: comparar el estadístico con el valor crítico, en lugar del p-valor con α. Llevan siempre a la misma decisión.

Bilateral o unilateral

Bilateral: $H_1: p \neq 0{,}5$

- Interesa un desvío en cualquier dirección.
- Región crítica: $|z| > 1{,}96$.
- Kerrich: p-valor = 0,1802.

Unilateral: $H_1: p > 0{,}5$

- Solo interesa una dirección, y se decide **antes** de ver los datos.
- Región crítica: $z > 1{,}645$.
- Kerrich: p-valor = 0,0901.

La unilateral tiene más potencia en su dirección y ninguna en la otra.

PREGUNTA PARA LA CLASE

Con la moneda de Kerrich no se rechazó Ho. ¿Quedó demostrado que la moneda es equilibrada?

4

Pruebas para una proporción y una media

Las dos pruebas de hoy

La prueba para una proporción

Hipótesis

H_0: $p = p_0$, contra $p \neq p_0$, $p > p_0$ o $p < p_0$.

Estadístico

$$z = (\hat{p} - p_0) / \sqrt{(p_0(1 - p_0) / n)}$$

Con p_0: la cuenta supone H_0 cierta.

Condición

$np_0 \geq 5$ y $n(1 - p_0) \geq 5$.

Para usar la normal.

La prueba t para una media

Los datos

La muestra de 30 hogares de la semana 9: $\bar{x}$ = 1,400, s = 1,003, EE = 0,183.

Las hipótesis

Un informe anterior decía 1,8 ocupados por hogar. H_0: μ = 1,8; H_1: $\mu \neq 1{,}8$.

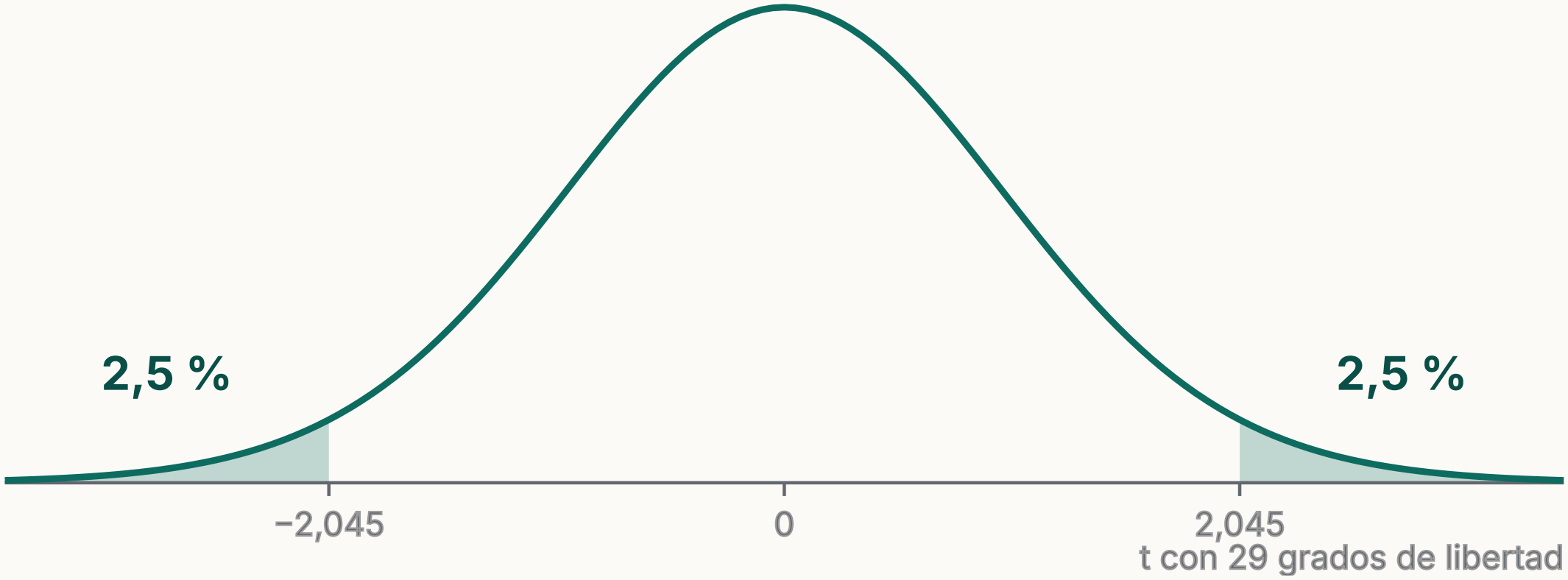
El estadístico

$t = (\bar{x} - \mu_0) / EE = (1{,}400 - 1{,}8) / 0{,}183 =$ **−2,19**

Con 29 grados de libertad.

La decisión

Valor crítico de la t con 29 grados de libertad: ±2,045. El t de la muestra, −2,19, cae en la región crítica.



Con la calculadora, p-valor = 0,0371: menor que 0,05. Se rechaza H_0.

La prueba y el intervalo dicen lo mismo

El intervalo del 95 % de la semana 9: de 1,026 a 1,774 ocupados por hogar.

Ho: μ = 1,8

1,8 queda afuera del intervalo: se rechaza. t = −2,19, p = 0,037.

Decisión correcta: μ no es 1,8.

Ho: μ = 1,6

1,6 queda adentro: no se rechaza. t = −1,09, p = 0,283.

Error de tipo II: μ = 1,471, no 1,6.

5

Leer bien un resultado

Significativo no quiere decir importante

Los pasos de una prueba

- Plantear H_0 y H_1, y elegir α, **antes** de ver los datos.
- Verificar las condiciones: muestra aleatoria, tamaño, forma de los datos.
- Calcular el estadístico de prueba.
- Calcular el p-valor, o comparar con el valor crítico.
- Decidir y redactar la conclusión **en el contexto del problema**, con el estadístico y el p-valor.

Significativo no es lo mismo que importante

Una moneda que sale cara el 50,1 % de las veces, tirada un millón de veces.

$z = 2{,}00$

$(0{,}501 - 0{,}5) / \sqrt{(0{,}25 / 1.000.000)}$

$p = 0{,}0456$

«estadísticamente significativo» con $\alpha = 0{,}05$

Con muestras enormes, cualquier diferencia, por mínima que sea, da un p-valor chico.

Qué es y qué no es el p-valor

Es

- La probabilidad de datos como los observados, o más extremos, **si H_0 fuera cierta**.
- Una medida de qué tan compatibles son los datos con H_0.

No es

- La probabilidad de que H_0 sea cierta.
- La probabilidad de que el resultado sea casualidad.
- Una medida del tamaño o la importancia del efecto.

Para llevarse

- Una prueba supone H_0 y pregunta qué tan raros serían los datos si fuera cierta.
- Hay dos errores: rechazar una H_0 cierta (α) y no rechazar una falsa (β).
- El p-valor es la probabilidad de un resultado así o más extremo si H_0 fuera cierta; se rechaza H_0 si $p \leq \alpha$.
- No rechazar H_0 no la demuestra; rechazarla no dice que el efecto sea importante.
- La prueba bilateral y el intervalo de confianza llevan a la misma decisión.

Guía de trabajo práctico 10
estadistica.ar/docentes/trabajos-practicos/semana-10

Simulador de errores de tipo I y II
estadistica.ar/herramientas/sim-errores-i-ii

Simulador del p-valor
estadistica.ar/herramientas/sim-p-valor

Calculadora de prueba t
estadistica.ar/herramientas/calc-prueba-t