Test o prueba de hipótesis

Agenda

Tests de hipótesis

Tests construidos a partir de pivotes

Motivación

Frecuentemente se desea decidir, en base a la información que provee la muestra, entre dos hipótesis relativas al parámetro de interés.

Los métodos para lograr esto constituyen la parte de la inferencia estadística llamada **pruebas o tests de hipótesis**.

El procedimiento formal para test de hipótesis es semejante al método científico.

Observar la naturaleza, formular una teoría y confrontarla con lo observado.

En el contexto de la estadística:

- tenemos una hipótesis sobre la población (por ejemplo acerca de la media, de la varianza de la población o la proporción de individuos con cierta característica)
- queremos saber si esa hipótesis es cierta o no
- tomamos una muestra, y en función de la muestra observada, decidimos si rechazamos o no esa hipótesis.

Ejemplo

Supongamos que un equipo de desarrollo está evaluando un nuevo sistema de bases de datos que, según el fabricante, tiene una **latencia promedio menor o igual a** 25 *ms* **por consulta.**

Para validar esta afirmación, el equipo realiza una **muestra aleatoria de 20 consultas típicas** y mide el tiempo de respuesta del sistema.

Utilizarán estos resultados para determinar si aceptan o no la hipótesis del fabricante de que la latencia media poblacional es menor o igual a 25 *ms* por consulta.

Hipótesis Estadística

Una **hipótesis estadística** suele ser una afirmación o conjetura sobre un conjunto de parámetros de una distribución de población o sobre la distribución en sí.

Se llama hipótesis porque no se sabe si es verdad o no.

El objetivo es desarrollar un procedimiento para determinar si los valores de una muestra observada de la población son consistentes con la hipótesis.

Si se considera que la muestra observada no es consistente con la hipótesis, decimos que la misma es *rechazada*; de lo contrario, decimos que no es rechazada.

Ejemplo (continuación)

Consideremos que en el sistema de bases de datos del ejemplo el tiempo de latencia en ms tiene distribución $N(\mu, 4)$ (varianza conocida de 4 ms^2).

La afirmación "μ es menor o igual que 25 ms" es nuestra hipótesis estadística.

Si una muestra de 16 consultas tiene un valor promedio de 25.2 ms, entonces, aunque este resultado no puede considerarse como evidencia a favor de la hipótesis "$\mu \leq 25$ ms", no es inconsistente con esta hipótesis, que por lo tanto no sería rechazada.

Ejemplo (continuación)

Pero, si la muestra de 16 consultas tiene un valor promedio de 32.5 ms, entonces a pesar de que una latencia tan alta es posible cuando $\mu \leq 25$ ms, es tan improbable que parece inconsistente con esta hipótesis, por lo que la misma sería rechazada.

Interpretación

Con una latencia promedio de 32.5 ms, parece haber evidencia suficiente para rechazar la hipótesis de que el sistema cumple con el requisito de que el tiempo medio de latencia es de 25 ms o menos.

Falta de evidencia

Debe tenerse en cuenta que el no rechazo de una hipótesis no supone ninguna garantía de la certeza de ésta, sino la falta de evidencia en contra de su veracidad.

Se podría considerar cierto parecido con la situación de una persona que está siendo juzgada según el principio de presunción de inocencia, de forma que sólo se rechaza su inocencia, es decir, la hipótesis, en caso de encontrar pruebas suficientes en su contra.

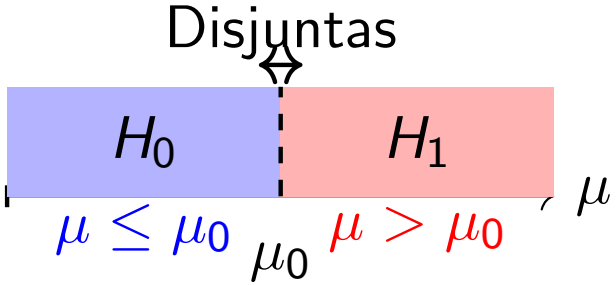
Clases de Hipótesis

- La hipótesis sobre la cual se estructura el test se denomina **Hipótesis Nula**, y es denotada por H_0,
- La hipótesis elegida como contraste a la hipótesis nula se denomina **Hipótesis Alternativa** y se denota por H_1.

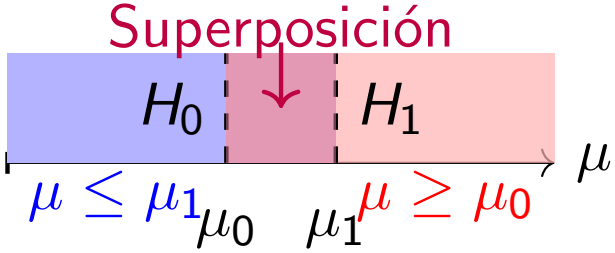
Condición para hipótesis válidas

Las hipótesis nula (H_0) y alternativa (H_1) deben definir regiones **disjuntas** para el parámetro μ.

Hipótesis válidas



Hipótesis inválidas



Tipos de tests en el contexto de latencia

Clasificación de tests

- **Unilaterales**: alternativa es de la forma $H_1 : \mu > 25$ ms o $H_1 : \mu < 25$ ms
- **Bilaterales**: $H_0 : \mu = 25$ $H_1 : \mu \neq 25$ ms

Tenemos dos posibles hipótesis nulas

Hipótesis Simple

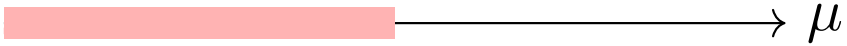
Punto único

$H_0 : \mu = 25$

Distribución completamente especificada: $N(25, 4)$

Hipótesis Compuesta

Intervalo



$H_0 : \mu \leq 25$

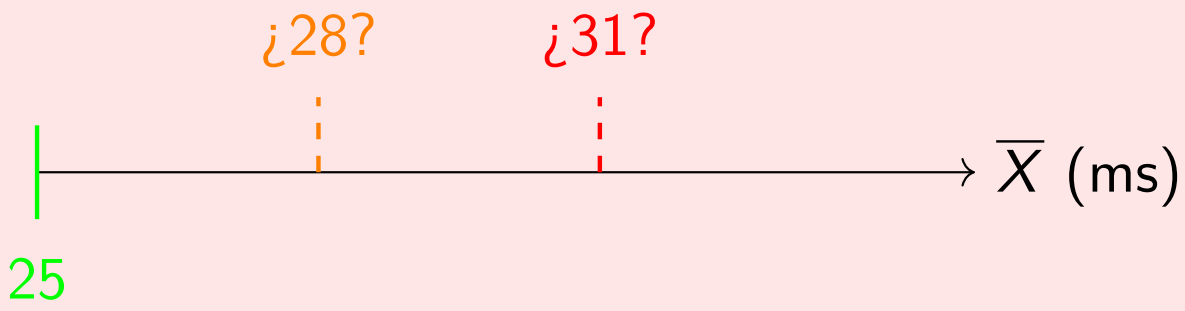
Familia de distribuciones: $N(\mu, 4)$ con $\mu \leq 25$

¿Como definir una regla de rechazo?

Si la muestra de tamaño 16 tiene un valor promedio de 32.5, dijimos que la hipótesis "$\mu \leq 25$" sería rechazada.

Pregunta clave

Con una muestra de 16 consultas: ¿A partir de qué valor de latencia muestral $\overline{X}$ deberíamos rechazar $H_0 : \mu \leq 25$ ms?



Necesito un criterio para decidir

Regiones de un test

Un test de hipótesis define una regla de decisión mediante una partición del espacio muestral en dos regiones,

- **región de aceptación**
- **región crítica o de rechazo**

Si la muestra considerada se encuentra dentro de la región crítica se rechaza la Hipótesis Nula, mientras que en el caso contrario no se rechaza dicha hipótesis al no existir evidencias para rechazarla.

Una forma de construir tales regiones es en función de un estadístico, llamado **estadístico del test**.

Ejemplo: Test unilateral para latencia del sistema

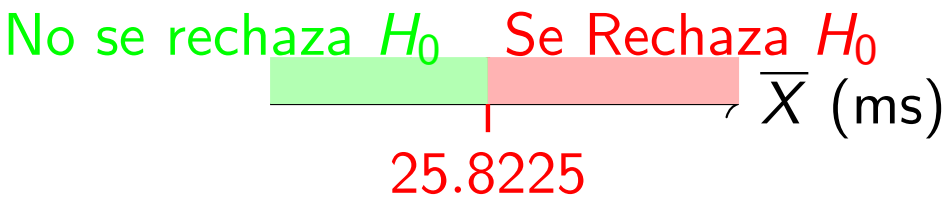
Contexto: Sistema de bases de datos con tiempo de latencia (en ms) dado por $X \sim N(\mu, 4)$, se realizan $n = 16$ consultas

Hipótesis

- $H_0 : \mu \leq 25$ ms (el sistema cumple con el estándar)
- $H_1 : \mu > 25$ ms (la latencia es mayor al estándar)

Un test popular tiene la siguiente **Región crítica** $(25.8225, +\infty)$:

$$\text{Se rechaza } H_0 \text{ si: } \overline{X} > 25.8225 \text{ ms}$$



Errores en tests de hipótesis

En cualquier test, pueden resultar dos tipos diferentes de errores:

- El **error de tipo I**, ocurre si se toma la decisión de rechazar H_0 cuando es cierta.
- El **error de tipo II**, ocurre si se toma la decisión de no rechazar H_0 cuando es falsa.

	Decisión estadística	
	No rechazar H_0	Rechazar H_0
H_0 verdadera	Correcta	Error tipo I
H_0 falsa	Error tipo II	Correcta

Error de tipo I

H_0 solo debe rechazarse si los datos resultantes son muy poco probables cuando H_0 es verdadera.

La forma clásica de lograr esto es especificar un valor α (chico) y luego requerir que el test tenga la propiedad de que siempre que H_0 sea verdadera, su probabilidad de ser rechazada nunca es mayor que α.

Es decir,

$$P[\text{error de tipo I}] \leq \alpha.$$

Cuando H_0 es simple,

$$P[\text{error de tipo I}] = \alpha.$$

Nivel de significación del test

El valor α, llamado **nivel de significación del test**, generalmente se establece de antemano, siendo algunos de los valores comúnmente elegidos $\alpha = 0.1, 0.05, 0.005$.

El enfoque clásico para testear H_0 es fijar un nivel de significancia α y luego requerir que la prueba tenga la propiedad de que la probabilidad de que ocurra un error tipo I nunca puede ser mayor que α.

Error de tipo II

La probabilidad de **error de tipo II** es denotado por la letra β.

Para cada situación particular indicativa de la falsedad de la hipótesis nula, se puede generar una variedad de errores de tipo II.

Esto puede dificultar su cálculo debido a que la probabilidad de no rechazar la hipótesis nula se calcula bajo la consideración de que la hipótesis nula es falsa.

Nivel del test: Cálculo y interpretación

Ejemplo del sistema de bases de datos el tiempo de latencia (en ms) está dado por $X \sim N(\mu, 4)$, se realizan $n = 16$ consultas

Hipótesis

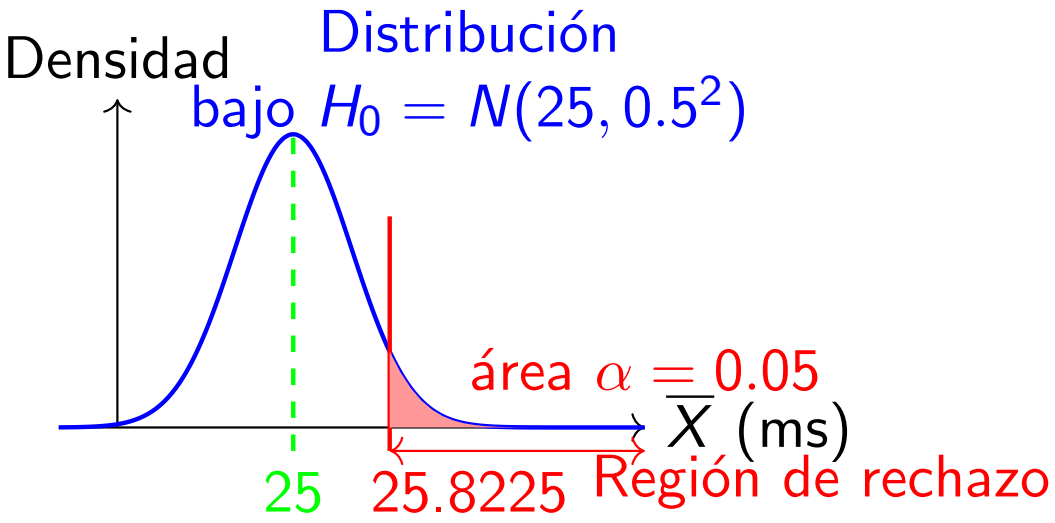
- $H_0 : \mu = 25$ ms (el sistema cumple con el estándar)
- $H_1 : \mu > 25$ ms (la latencia es mayor al estándar)

Se rechaza H_0 si: $\overline{X} > 25.8225$ ms

Distribución bajo H_0

$$\overline{X} \sim N(25, \frac{4}{16}) \sim N(25, 0.5^2)$$

$$\begin{aligned}\alpha &= P(\overline{X} > 25.8225) \\ &= P\left(Z > \frac{25.8225 - 25}{0.5}\right) \\ &= P(Z > 1.645) = 1 - \Phi(1.645) = 0.05\end{aligned}$$



Significado práctico

Si el sistema realmente tiene latencia media de 25 ms, solo un 5 % de las muestras que obtengamos tendrán media muestral mayor a 25.82 ms y producirán que rechacemos incorrectamente H_0.

Metodología: Tests mediante pivotes - Caso normal

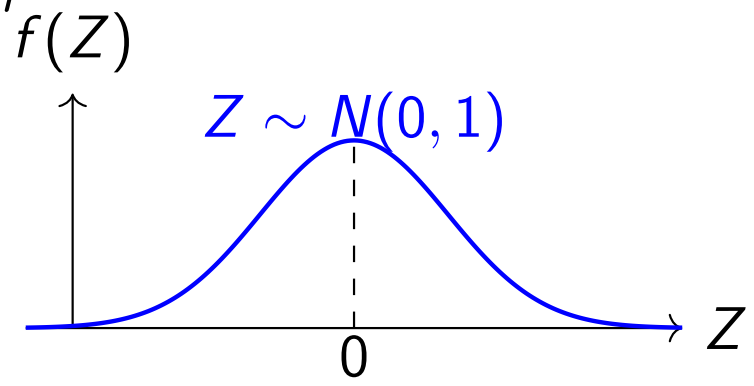
Contexto: $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ con σ^2 conocida

Pivote

Para la media μ, el pivote natural es:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Tiene **distribución conocida**: $N(0,1)$ independiente de μ y podemos despejar μ en términos de Z.



Construcción del test unilateral derecho

Hipótesis: $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$

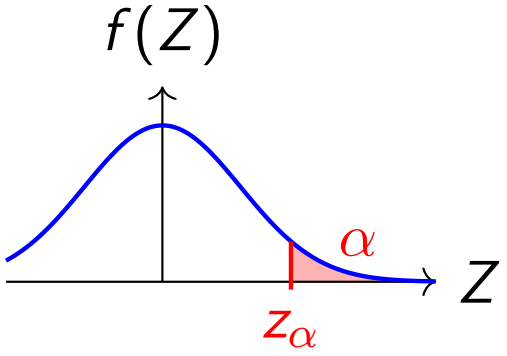
Paso 1: Definir el pivote bajo H_0 (**Estadístico de prueba**)

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{bajo } H_0$$

Paso 2: Dado α, construir la región de rechazo

Rechazar si:

$$\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$$



Construcción del test unilateral izquierdo

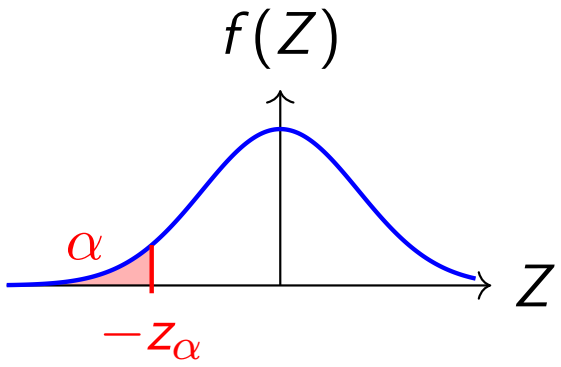
Hipótesis: $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$

Paso 1: Definir el Estadístico de prueba

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{bajo } H_0$$

Paso 2: Dado α, construir la región de rechazo

Rechazar si: $$\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha$$



Construcción del test bilateral

Hipótesis: $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

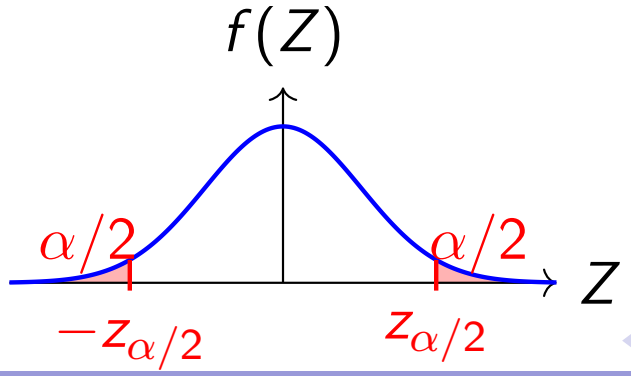
Paso 1: Definir el Estadístico de prueba

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{bajo } H_0$$

Paso 2: Dado α, construir la región de rechazo

Rechazar H_0 si $|Z|$ es "grande":

$$|Z| > z_{\alpha/2}$$



Ejemplo aplicado: Sistema de bases de datos

Ejemplo del sistema de bases de datos el tiempo de latencia (en ms) está dado por $X \sim N(\mu, 4)$, se realizan $n = 16$ consultas y se obtiene $\overline{x} = 26$.

Hipótesis

- $H_0 : \mu = 25$ ms (el sistema cumple con el estándar)
- $H_1 : \mu > 25$ ms (la latencia es mayor al estándar)

Paso 1: Identificar el estadístico de prueba

$$Z = \frac{\overline{X} - 25}{2/\sqrt{16}} = \frac{\overline{X} - 25}{0.5} \sim N(0, 1) \text{ bajo } H_0$$

Paso 2: Región de rechazo

- Cuantil: $z_{0.05} = 1.645$
- Región de rechazo: $Z = \frac{\overline{X}-25}{0.5} > 1.645$

Reemplazamos $\overline{X}$ por $\overline{x} = 26$

$$\frac{26-25}{0.5} = 2 > 1.645$$

Por lo tanto rechazamos H_0 con un nivel de significación 0.05. Es decir, el tiempo de latencia medio es mayor a 25 ms.