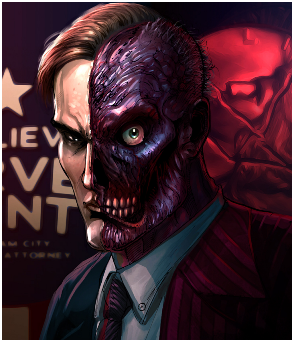
MAT-206: Test de hipótesis I

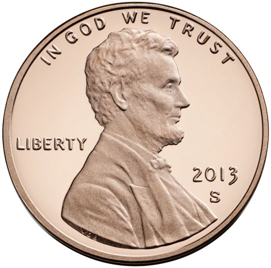
Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Decidir lanzando una moneda[1]





[1] Harvey Dent, o “Dos Caras”, un enemigo de Batman.

Test de hipótesis

Suponga un modelo estadistico asociado a la muestra $\boldsymbol{X} = (X_1, \dots, X_n)^\top$, como:

$$\mathcal{P} = \{\mathsf{P}_\theta : \boldsymbol{\theta} \in \Theta\}.$$

De este modo, hipótesis son subconjuntos de $\mathcal{P}$ formulados mediante particionar el espacio paramétrico

$$\Theta = \Theta_0 \cup \Theta_1, \qquad \Theta_0 \cap \Theta_1 = \varnothing$$

Esta configuración es habitual en el lenguaje cienfífico, donde es usual contar con dos hipótesis que se desean comparar. La hipótesis nula H_0, establece que $\theta \in \Theta_0$,

$$H_0 : \theta \in \Theta_0,$$

mientras que la hipótesis alternativa postula que $\theta \in \Theta_1$, es decir

$$H_1 : \theta \in \Theta_1.$$

Test de hipótesis

Es decir, basado en los datos observados, deseamos decidir entre los modelos:

$$\mathcal{P}_0 = \{\mathsf{P}_\theta : \boldsymbol{\theta} \in \Theta_0\}, \quad \text{versus} \qquad \mathcal{P}_1 = \{\mathsf{P}_\theta : \boldsymbol{\theta} \in \Theta_1\}.$$

Decimos que la hipótesis H_0 es verdadera si existe evidencia de que la muestra es distribuída de acuerdo a $\mathsf{P}_\theta \in \mathcal{P}_0$.

Una hipótesis es llamada simple si especifíca completamente la distribución de interés, por ejemplo:

$$H_0 : \theta = \theta_0, \quad \text{versus} \qquad H_1 : \theta = \theta_1.$$

En caso contrario se dice hipotesis compuesta,

$$H_0 : \theta > \theta_0, \quad \text{versus} \qquad H_1 : \theta \leq \theta_1.$$

Además, hipotesis del tipo:

$$H_0 : \theta = \theta_0, \quad \text{versus} \qquad H_1 : \theta \neq \theta_1,$$

se dicen a dos colas.

Definición 1:

Un test δ es cualquier función $\delta : \mathcal{X}^n \to \{0, 1\}$.

Una función δ adopta el valor 0 cuando decidimos en favor de H_0, en caso contrario toma el valor 1. Es decir, cuando decidimos en favor de H_1.

Usualmente escribimos un test, como:

$$\delta(X_1, \dots, X_n) = \begin{cases} 1, & T(X_1, \dots, X_n) \in C \\ 0, & T(X_1, \dots, X_n) \notin C \end{cases},$$

donde T es llamado estadístico de prueba y C denota la región de rechazo.

¿Cómo luce una regla de decisión?

Considere el **criterio** para aprobar una asignatura.

Suponga que usted ha obtenido las siguientes notas en (por ejemplo) MAT-206:

$$\boldsymbol{x} = \{x_1, x_2, x_3\} = \{68, 32, 70\}.$$

De este modo, el profesor calcula su promedio obteniendo: $\overline{x} = 56$. Por tanto,

$$\overline{x} \geq 55, \qquad \text{es decir, Ud. ha aprobado.}$$

Podemos reescribir lo anterior como la siguiente regla de decisión:

- Si $\overline{x} \in [55, 100]$, el alumno es aprobado.
- En caso contrario,[2] el alumno reprueba la asignatura.

[2] Es decir, $\overline{x} \notin [55, 100]$

Observación:

Un test, descansa en la elección de T y C. Note que δ es una variable aleatoria Bernoulli. En efecto,

$$\delta = \begin{cases} 1, & \text{con probabilidad } \mathsf{P}(T(X_1,\dots,X_n) \in C), \\ 0, & \text{con probabilidad } \mathsf{P}(T(X_1,\dots,X_n) \notin C). \end{cases}$$

¿Cómo escoger T y C tal que δ sea una buena función test?

Cuando $H_0 : \theta \in \Theta_0$ es verdad esperamos que la distribución de $\delta(X_1,\dots,X_n)$ se encuentre concentrada en torno de 0. En caso contrario, es decir, si $H_1 : \theta \in \Theta_1$ es verdadera, $\delta(X_1,\dots,X_n)$ estará concentrada en torno de 1.

Un poco de controversia: Resultado de un juicio penal

Parecer de un juez:

¿La evidencia dada por el fiscal es suficiente para declarar culpable al acusado?

El juez tiene 2 opciones:

- Declarar al acusado culpable.
- Declarar al acusado inocente.

En términos científicos, debemos plantear las siguientes hipótesis:

$$H_0 : \text{el acusado es inocente.}$$
$$H_1 : \text{el acusado es culpable.}$$

Así, el fiscal debe probar que el acusado es culpable, *más allá de toda duda razonable*.

EX UMBRA IN SOLEM

Test de hipótesis

Un poco de controversia: Resultado de un juicio penal

Parecer de un juez:

¿La evidencia dada por el fiscal es suficiente para declarar culpable al acusado?

El juez tiene 2 opciones:

- Declarar al acusado culpable.
- Declarar al acusado inocente.

En términos científicos, debemos plantear las siguientes hipótesis:

$$H_0 : \text{el acusado es inocente.}$$
$$H_1 : \text{el acusado es culpable.}$$

Así, el fiscal debe probar que el acusado es culpable, *más allá de toda duda razonable*.

Test de hipótesis

De este modo, en el juicio puede ocurrir lo siguiente:

Decisión	El acusado es	
	inocente	culpable
preso	falso positivo	OK
libre	OK	falso negativo

En la nomenclatura de test de hipótesis, tenemos:

Decisión	El acusado es	
	H_0 es verdadero	H_1 es verdadero
rechazar H_0	error tipo I	OK
aceptar H_0	OK	error tipo II

Test de hipótesis

De este modo, en el juicio puede ocurrir lo siguiente:

Decisión	El acusado es	
	inocente	culpable
preso	falso positivo	OK
libre	OK	falso negativo

En la nomenclatura de test de hipótesis, tenemos:

Decisión	El acusado es	
	H_0 es verdadero	H_1 es verdadero
rechazar H_0	error tipo I	OK
aceptar H_0	OK	error tipo II

Definición 2 (Probabilidad de error):

Sea $H_0 : \theta \in \Theta_0$ y $H_1 : \theta \in \Theta_1$, dos hipótesis de interés. La probabilidad de error de tipo I se define como la función $h : \Theta_0 \to [0,1]$,

$$h(\theta) = \mathsf{P}_\theta(\delta = 1), \qquad \theta \in \Theta_0,$$

mientras que la probabilidad de error de tipo II es definido como la función $g : \Theta_1 \to [0,1]$

$$g(\theta) = \mathsf{P}_\theta(\delta = 0), \qquad \theta \in \Theta_1.$$

Observación:

Evidentemente, $h(\theta) \neq 1 - g(\theta)$.

Observación:

La idea es escoger T y C tal que las probabilidades de error de tipo I y II sean lo más pequeñas posibles.

Ejercicio:

Sea $X_1, \dots, X_n$ variables aleatorias IID desde $\mathsf{N}(\mu, 1)$. Deseamos probar la hipótesis

$$H_0 : \mu = 0, \qquad \text{versus} \qquad H_1 : \mu \neq 0,$$

usando la estadística de prueba

$$T_n(X_1, \dots, X_n) = \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

y el test

$$\delta(X_1, \dots, X_n) = \begin{cases} 1, & \text{si } |T_n(X_1, \dots, X_n)| \geq Q, \\ 0, & \text{en otro caso} \end{cases},$$

donde $Q > 0$.

(a) Determine la probabilidad de error de tipo I.

(b) Halle la probabilidad de cometer un error de tipo II.

Test de hipótesis

Sea

$$\delta(X_1,\dots,X_n) = I\{T(X_1,\dots,X_n) \in C\},$$

un test y suponga que deseamos reducir su probabilidad de error de tipo I,

$$h(\theta) = \mathsf{P}_\theta(\delta = 1), \qquad \theta \in \Theta_0,$$

sobre todo $\theta \in \Theta_0$. Para esto debemos "rechazar menos frecuentemente", es decir, reemplazar C por $C_* \subset C$ y obtener

$$\delta_* = I\{T(X_1,\dots,X_n) \in C_*\}.$$

Note que

$$\begin{aligned}\mathsf{P}_\theta(\delta_* = 1) &= \mathsf{P}(T(X_1,\dots,X_n) \in C_*) \leq \mathsf{P}(T(X_1,\dots,X_n) \in C) \\ &= \mathsf{P}_\theta(\delta = 1), \quad \forall\, \theta \in \Theta_0.\end{aligned}$$

Mientras que $C_* \subset C \Rightarrow C_*^c \supset C^c$, y de este modo,

$$\begin{aligned}\mathsf{P}_\theta(\delta_* = 0) &= \mathsf{P}(T(X_1,\dots,X_n) \notin C_*) \geq \mathsf{P}(T(X_1,\dots,X_n) \notin C) \\ &= \mathsf{P}_\theta(\delta = 0), \quad \forall\, \theta \in \Theta_1.\end{aligned}$$

Es decir, al intentar reducir el error de tipo I, se incrementa el error de tipo II!

Definición 3 (Método de Neyman-Pearson):

Sea $H_0 : \theta \in \Theta_0$ y $H_1 : \theta \in \Theta_1$ dos hipótesis de interés.

1. Fijar $\alpha \in (0, 1)$ el nivel de significancia o tamaño del test.
2. Considere test $\delta : \mathcal{X}^n \to \{0, 1\}$ tales que

$$\sup_{\theta \in \Theta_0} \mathsf{P}_\theta(\delta = 1) \leq \alpha,$$

   es decir, nos restringiremos a la clase

$$\mathcal{D}(\Theta_0, \alpha) = \Big\{\delta : \mathcal{X}^n \to \{0, 1\} \Big| \sup_{\theta \in \Theta_0} \mathsf{P}_\theta(\delta = 1) \leq \alpha\Big\}.$$

3. En la clase $\mathcal{D}(\Theta_0, \alpha)$ considere test con menor probabilidad de error tipo II

$$g(\theta) = \mathsf{P}_\theta(\delta = 0), \quad \theta \in \Theta_1,$$

   o equivalentemente, aquél test con mayor poder

$$\beta(\theta) = 1 - g(\theta) = \mathsf{P}_\theta(\delta = 1), \quad \theta \in \Theta_1.$$

Test de hipótesis

A continuación revisamos métodos para construir tests. Por simplicidad consideramos el caso 1-dimensional y los siguientes tipos de hipótesis.

1. Simple vs. Simple:

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta = \theta_1$$

2. Unilateral vs. Unilateral:

$$H_0 : \theta \leq \theta_0 \qquad \text{versus} \qquad H_1 : \theta > \theta_1$$
$$H_0 : \theta \geq \theta_0 \qquad \text{versus} \qquad H_1 : \theta < \theta_1$$

3. Simple vs. Bilateral:

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta \neq \theta_1$$

Definición 4 (Test optimal):

Un test δ de $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ es llamado óptimo[3] de nivel α, si satisface que:

1. $\delta \in \mathcal{D}(\Theta_0, \alpha)$.
2. $\mathsf{P}_{\theta_1}(\psi = 1) \leq \mathsf{P}_{\theta_1}(\delta = 1)\ \forall\, \theta_1 \in \Theta_1,\ \forall\, \psi \in \mathcal{D}(\Theta_0, \alpha)$.

[3] o uniformemente más poderoso de nivel α

Lema 1 (Neyman-Pearson), Caso simple:

Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ con densidad conjunta $f(\boldsymbol{x}; \theta)$ y suponga que deseamos probar

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta = \theta_1,$$

para algún nivel $\alpha \in (0, 1)$, para $\theta_0 \neq \theta_1$. Si la variable aleatoria

$$\Lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x}; \theta_1)}{f(\boldsymbol{x}; \theta_0)} = \frac{L(\theta_1)}{L(\theta_0)},$$

es tal que existe un $Q > 0$ satisfaciendo

$$\mathsf{P}_{\theta_0}(\Lambda > Q) = \alpha,$$

entonces el test dado por

$$\delta(\boldsymbol{X}) = I\{\Lambda(\boldsymbol{X}) > Q\},$$

es un test óptimo para probar H_0 versus H_1 a un nivel α.

Ejemplo:

Sea $X_1, \dots, X_n$ variables IID desde Exp(λ) y sea $\lambda_1 > \lambda_0$ dos constantes y considere el problema:

$$H_0 : \lambda = \lambda_0, \qquad \text{versus} \qquad H_1 : \lambda = \lambda_1.$$

La verosimilitud es

$$f(\boldsymbol{x}; \lambda) = \prod_{i=1}^n \lambda \exp(-\lambda x_i) = \lambda^n \exp\Big(-\lambda \sum_{i=1}^n x_i\Big).$$

De acuerdo con el Lema de Neyman-Pearson, tenemos

$$\Lambda(X_1, \dots, X_n) = \frac{f(\boldsymbol{X}; \lambda_1)}{f(\boldsymbol{X}; \lambda_0)} = \Big(\frac{\lambda_1}{\lambda_0}\Big)^n \exp\Big[(\lambda_0 - \lambda_1) \sum_{i=1}^n X_i\Big],$$

y se rechaza la hipótesis nula si $\Lambda > Q$, para Q tal que

$$\mathsf{P}_{\lambda_0}(\Lambda(X_1, \dots, X_n) > Q) = \alpha,$$

siempre que Q exista.

Note que $\Lambda(\boldsymbol{X})$ es función decreciente de $T(X_1,\dots,X_n) = \sum_{i=1}^n X_i$, pues $\lambda_0 < \lambda_1$. Así

$$\Lambda(X_1,\dots,X_n) \geq Q \quad \Longleftrightarrow \quad T(X_1,\dots,X_n) \leq q,$$

para algún q satisfaciendo

$$\alpha = \mathsf{P}_{\lambda_0}(\Lambda(\boldsymbol{X}) \geq Q) \quad \Longleftrightarrow \quad \alpha = \mathsf{P}_{\lambda_0}(T(\boldsymbol{X}) \leq q).$$

Bajo H_0, $T(\boldsymbol{X}) \sim \mathsf{Gama}(n,\lambda_0)$. De ahí que q es tal que

$$\alpha = \mathsf{P}_{\lambda_0}(T(\boldsymbol{X}) \leq q),$$

es decir q_α es un cuantil de la distribución $\mathsf{Gama}(n,\lambda_0)$.

Ejemplo:

Sea $X_1, \dots, X_n \sim$ FE(θ) 1-paramétrica, donde

$$f(x;\theta) = \exp\{\eta(\theta)T(x) - b(\theta)\}h(x),$$

con η creciente. Suponga que deseamos probar:

$$H_0 : \theta = \theta_0, \qquad \text{contra} \qquad H_1 : \theta = \theta_1.$$

Además, asumiremos que $\theta_0 < \theta_1$. El Lema de Neyman-Pearson lleva al estadístico de prueba

$$\delta = I\{L(\theta_1)/L(\theta_0) > Q\} = I\{\log L(\theta_1) - \log L(\theta_0) > \log Q\}.$$

Para el caso de la familia exponencial, tenemos

$$\begin{aligned}
\delta &= I\Big\{(\eta(\theta_1) - \eta(\theta_0)) \sum_{i=1}^n T(X_i) - n(b(\theta_1) - b(\theta_0)) > \log Q\Big\} \\
&= I\Big\{\sum_{i=1}^n T(X_i) > \frac{\log Q + n(b(\theta_1) - b(\theta_0))}{\eta(\theta_1) - \eta(\theta_0)}\Big\}.
\end{aligned}$$

Como η es creciente, $\eta(\theta_1) - \eta(\theta_0) > 0$ y $n(b(\theta_1) - b(\theta_0))$ es solo una constante. De modo que podemos escribir

$$\delta = I\{T(X_1, \dots, X_n) > q\}.$$

Si T es una variable aleatoria continua, y deseamos un test de nivel α, entonces q es el cuantil $(1-\alpha)$ de

$$G_0(t) = \mathsf{P}_{\theta_0}(T(X_1, \dots, X_n) \leq t),$$

es decir, corresponde al cuantil-$(1-\alpha)$ de la distribución de muestreo de $T(\boldsymbol{X})$ cuando el parámetro es θ_0 (es decir la distribución nula de T).

Observación:

Cuando η es decreciente, entonces para $\theta_0 < \theta_1$ tenemos $\eta(\theta_1) - \eta(\theta_0) < 0$ y el test óptimo resulta:

$$\delta = I\{T(X_1, \dots, X_n) \leq q\}.$$

Si deseamos un test de tamaño α y q debe ser un cuantíl-α de

$$G_0(t) = \mathsf{P}_{\theta_0}(T(X_1, \dots, X_n) \leq t).$$

Es decir, rechazamos $H_0 : \theta = \theta_0$ (en favor de $H_1 : \theta = \theta_1$) si:

$$T(X_1, \dots, X_n) \leq q_\alpha,$$

con q_α el valor cuantil de nivel α de la distribución $G_0(t) = \mathsf{P}_{\theta_0}(T(\boldsymbol{X}) \leq t)$.

EX UMBRA IN SOLEM

Para la hipótesis

$$H_0 : \theta = \theta_0, \qquad \text{versus} \qquad H_1 : \theta = \theta_1,$$

y sea $X_1, \ldots, X_n \sim \mathsf{FE}(\theta)$ 1-paramétrica. Tenemos la siguiente tabla:

η	$\theta_0 < \theta_1$	$\theta_0 > \theta_1$
creciente	$I\{T(X_1, \ldots, X_n) > q_{1-\alpha}\}$	$I\{T(X_1, \ldots, X_n) \leq q_\alpha\}$
decreciente	$I\{T(X_1, \ldots, X_n) \leq q_\alpha\}$	$I\{T(X_1, \ldots, X_n) > q_{1-\alpha}\}$