CONTRASTE DE HIPOTESIS

Julián de la Horra
Departamento de Matemáticas U.A.M.

1 Introducción

En este capítulo, vamos a abordar el Contraste de Hipótesis, que es el tercer y último gran conjunto de técnicas que se utilizan en la Inferencia Estadística. La situación general que vamos a considerar es la misma que en los capítulos anteriores:

Disponemos de una muestra aleatoria $(X_1, ..., X_n)$ de una característica X de una población. Pensamos que esta característica puede ser adecuadamente modelizada mediante un modelo de probabilidad con función de masa $P_\theta(x)$ (en el caso discreto) o con función de densidad $f_\theta(x)$ (en el caso continuo). En cualquiera de los casos, lo único que nos falta por conocer es el valor del parámetro $\theta \in \Theta$, que es desconocido.

En este capítulo, no tratamos de dar una estimación de este parámetro desconocido, sino de decidir, a la vista de la información muestral, si aceptamos o rechazamos que este parámetro tiene algún valor dentro de una región del espacio paramétrico. Un par de ejemplos sencillos ayudarán a entender el tipo de problemas que se pueden abordar y resolver mediante los contrastes de hipótesis.

Ejemplo 1.- En los ejercicios de cálculo de probabilidades, siempre se suele hablar de monedas equilibradas pero, naturalmente, no todas lo son. Nos gustaría decidir si es razonable aceptar que una determinada moneda es (aproximadamente) equilibrada, es decir, nos gustaría decidir si es razonable aceptar que su probabilidad de cara es (aproximadamente) 1/2. Llamaremos $p = P(Cara)$.

Necesitamos datos, para lo cual lanzamos la moneda, por ejemplo, 100 veces, y anotamos los resultados.

Desde un punto de vista formal, las caras y las cruces pueden ser codificadas mediante unos y ceros, de modo que tenemos una muestra aleatoria $(X_1, ..., X_{100})$ de

$$X = \begin{cases} 1 \quad \text{(si sale cara)} & \text{con probabilidad } p \\ 0 \quad \text{(si sale cruz)} & \text{con probabilidad } 1-p \end{cases}$$

y, por tanto, X puede ser modelizada mediante un modelo de Bernoulli con parámetro p desoconocido.

La pregunta a la que tenemos que dar respuesta es la siguiente: ¿Podemos aceptar, a la vista de los resultados muestrales, que la probabilidad de cara es (aproximadamente) 1/2?

La decisión la tomaremos en función del número de caras obtenidas. En algunos casos, esta decisión es intuitivamente clara:

Si obtenemos 50 caras (y 50 cruces), obviamente aceptaremos que la moneda está equilibrada.

Si obtenemos 95 caras (y 5 cruces), obviamente rechazaremos que la moneda está equilibrada.

Sin embargo, si obtenemos 65 caras (y 35 cruces), es posible que haya división de opiniones sobre si la moneda está equilibrada.

Para decidir qué hacemos, sobre todo en los casos conflictivos, necesitamos una metodología general que nos permita resolver este tipo de problemas de un modo sistemático y lo más objetivo posible.

Ejemplo 2.- En una fábrica, se está ensayando una nueva fibra sintética, y se quiere tener una razonable seguridad de que la resistencia media a la rotura de las cuerdas fabricadas con esta nueva fibra es superior a 30 unidades, para estar seguros de satisfacer las especificaciones requeridas por la Agencia de Control de Calidad. Llamaremos μ al valor de la resistencia media de toda la producción, que es desconocida.

Necesitamos datos, para lo cual medimos la resistencia de, por ejemplo, 100 cuerdas, y anotamos los resultados.

Desde un punto de vista formal, lo que tenemos es una muestra aleatoria $(X_1, ..., X_{100})$ de la característica $X =$ "Resistencia a la rotura", que puede ser modelizada mediante una distribución $N(\mu; \sigma)$, con parámetros μ y σ desconocidos.

La pregunta a la que tenemos que dar respuesta es la siguiente: ¿Podemos asegurar, a la vista de los resultados muestrales, que la resistencia media de toda la producción, μ, es superior a 30 unidades?

La decisión la tomaremos en función de la resistencia media que hayamos obtenido en la muestra. En algunos casos, esta decisión es intuitivamente clara:

Si la resistencia media muestral obtenida es 40, obviamente podemos asegurar que la resistencia media de toda la producción será superior a 30.

Si la resistencia media muestral obtenida es 28, obviamente no podemos asegurar que la resistencia media de toda la producción será superior a 30.

Sin embargo, si la resistencia media muestral obtenida es 31, es posible que haya división de opiniones sobre si estamos en condiciones de asegurar que la resistencia media de toda la producción será superior a 30.

Para decidir qué hacemos, sobre todo en los casos conflictivos, necesitamos una metodología general que nos permita resolver este tipo de problemas de un modo sistemático y lo más objetivo posible.

2 Contraste de hipótesis

En primer lugar, vamos a definir lo que entenderemos por un contraste de hipótesis sobre un parámetro:

Definición.- Sea $(X_1, \ldots, X_n)$ una muestra aleatoria de una característica X de una población con función de masa $P_\theta(x)$ (caso discreto), o con función de densidad $f_\theta(x)$ (caso continuo), donde $\theta \in \Theta$ es desconocido. Dividimos el espacio paramétrico, Θ, en dos regiones: Θ_0 y $\Theta_1 = \Theta - \Theta_0$. Utilizaremos la siguiente notación:

$H_0 : \theta \in \Theta_0$ (hipótesis nula)
$H_1 : \theta \in \Theta_1$ (hipótesis alternativa)

Un *test para contrastar la hipótesis nula* $H_0 : \theta \in \Theta_0$ *frente a la hipótesis alternativa* $H_1 : \theta \in \Theta_1$, consiste en decidir, para cada posible muestra, si aceptamos o rechazamos la hipótesis nula H_0; por lo tanto, un test consistirá en dividir el espacio muestral $\mathcal{X}$ (conjunto de todas las posibles muestras) en dos regiones: una región crítica R o de rechazo de H_0 y una región $A = \mathcal{X} - R$ de aceptación de H_0.

Por lo tanto, un test estadístico queda perfectamente especificado, en cuanto hayamos definido la región R, que nos indica cuál es el conjunto de muestras que nos llevarían a rechazar la hipótesis nula.

Podemos volver ahora a los dos ejemplos de la Introducción:

Ejemplo 1 (continuación).- La pregunta a la que teníamos que dar respuesta era la siguiente: ¿Podemos aceptar, a la vista de los resultados muestrales, que la probabilidad de cara es (aproximadamente) 1/2? El contraste de hipótesis que plantearíamos para responder a esta pregunta sería el siguiente:

$H_0 : p = 1/2$ (hipótesis nula)
$H_1 : p \neq 1/2$ (hipótesis alternativa)

Una posible región de rechazo de la hipótesis nula sería la siguiente:

$$\begin{aligned} R &= \{\text{Número de caras } < 35 \text{ ó Número de caras } > 65\} \\ &= \left\{\sum x_i < 35 \text{ ó } \sum x_i > 65\right\} \end{aligned}$$

Por supuesto, esta región de rechazo es, simplemente, una posibilidad. De momento, queda sin resolver el problema de cómo elegir la región de rechazo de la hipótesis nula.

Ejemplo 2 (continuación).- La pregunta a la que teníamos que dar respuesta era la siguiente: ¿Podemos asegurar, a la vista de los resultados muestrales, que la resistencia media de toda la producción, μ, es superior a 30 unidades? El contraste de hipótesis que plantearíamos para responder a esta pregunta sería el siguiente:

$H_0 : \mu \leq 30$ (hipótesis nula)
$H_1 : \mu > 30$ (hipótesis alternativa)

Una posible región de rechazo de la hipótesis nula sería la siguiente:

$$R = \{\bar{x} > 31\}$$

La conclusión con esta región de rechazo sería: Podemos "asegurar" que la resistencia media de toda la producción, μ, es superior a 30 unidades, si $\bar{x} > 31$.

Por supuesto, esta región de rechazo es, simplemente, una posibilidad. De momento, queda sin resolver el problema de cómo elegir la región de rechazo de la hipótesis nula.

Con un test se pueden cometer dos tipos de errores:

Error de tipo I: rechazar H_0 cuando H_0 es cierta.
Error de tipo II: aceptar H_0 cuando H_0 es falsa.

Evidentemente, nos gustaría que la probabilidad de cometer ambos tipos de errores fuera baja. Sin embargo, la forma en que habitualmente se procede es la siguiente: se elige H_0 de modo que el error de tipo I sea el más grave y, en consecuencia, se intenta que la probabilidad de error de Tipo I sea siempre pequeña. Estas ideas se formalizan mediante las siguientes definiciones:

Definición.- Sea $(X_1, \ldots, X_n)$ una muestra aleatoria de una característica X de una población con función de masa $P_\theta(x)$ (caso discreto), o con función de densidad $f_\theta(x)$ (caso continuo), donde $\theta \in \Theta$ es desconocido. Queremos contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$. Consideramos un test definido mediante la región R (de rechazo de la hipótesis nula).

La *función de potencia del test* R se define como la función que a cada posible valor del parámetro θ le hace corresponder el valor:

$P_\theta(R) =$ "Probabilidad de rechazar H_0 cuando el parámetro es θ".

El *nivel de significación del test* R se define de la siguiente manera:

$\alpha = \max_{\theta\in\Theta_0} P_\theta(R)$
$=$ "Máxima probabilidad de rechazar H_0, cuando H_0 es cierta"
$=$ "Máxima probabilidad de cometer un error de tipo I".

Por lo tanto, el nivel de significación mide la probabilidad de cometer el error más grave y, por este motivo, el nivel de significación siempre se elige próximo a 0.

Los valores tradicionalmente elegidos para α son: 0,10, 0,05 y 0,01. El más habitual de todos es $\alpha = 0{,}05$.

Si el nivel de significación es demasiado próximo a 0, su probabilidad de error de Tipo I será bajísima, pero a costa de que la probabilidad de error de Tipo II aumente. Por este motivo, suele tomarse $\alpha = 0{,}05$, que representa un valor de compromiso.

3 Método de razón de verosimilitudes

En esta sección, abordamos la cuestión de cómo construir tests de hipótesis de un modo sistemático y lo más objetivo posible. El método habitualmente utilizado es el *método de razón de verosimilitudes.*

De manera esquemática, los pasos que hay que dar para obtener un test de hipótesis, es decir, una región R (de rechazo de la hipótesis nula), mediante el método de razón de verosimilitudes, son los siguientes:

1. Fijamos un nivel de significación α (próximo a 0).

2. Obtenemos la función de verosimilitud:

$$L(\theta) = P_\theta(x_1, ..., x_n) = P_\theta(x_1) \ldots P_\theta(x_n)$$

   Por supuesto, si estamos en un caso continuo, utilizaríamos la función de densidad del modelo utilizado.

3. Consideramos el siguiente estadístico que, por razones obvias, recibe el nombre de cociente o *razón de verosimilitudes*:

$$RV(x_1, ..., x_n) = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} \in (0, 1)$$

   (RV = Iniciales de *Razón de Verosimilitudes*)

   El mensaje intuitivo que recibimos de este estadístico es el siguiente:

   Si $RV(x_1, ..., x_n)$ es próximo a 0, el numerador es pequeño, la muestra es poco verosímil bajo H_0, y lo mejor que podemos hacer es rechazar H_0.

   Si $RV(x_1, ..., x_n)$ es próximo a 1, el numerador es grande, la muestra es muy verosímil bajo H_0, y lo mejor que podemos hacer es aceptar H_0.

4. Todo lo anterior nos lleva a definir la siguiente región, R, de rechazo de la hipótesis nula:

$$R = \left\{ (x_1, \ldots, x_n) : RV(x_1, ..., x_n) = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} < c \right\}$$

   donde c se obtiene de la condición:

$$\alpha = \max_{\theta \in \Theta_0} P_\theta(R)$$

La popularidad de estos tests se debe, muy especialmente, a que en los casos más habituales quedan reducidos a reglas muy sencillas. A continuación, aplicaremos los pasos generales que se acaban de indicar a un caso concreto:

Ejemplo.- Supongamos que $(X_1, \ldots, X_n)$ es una muestra aleatoria de una característica $X \sim N(\mu; \sigma)$, donde los dos parámetros son desconocidos, de modo que, en este caso, $\theta = (\mu; \sigma)$. Estamos interesados en hacer un contraste de hipótesis sobre el posible valor de μ. Concretamente, vamos a obtener el test de razón de verosimilitudes para contrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$, al nivel de significación α. Aplicamos los pasos del método:

1. Fijamos un nivel de significación α (próximo a 0).

2. Obtenemos la función de verosimilitud:

$$\begin{aligned} L(\theta) &= f_\theta(x_1, ..., x_n) = f_\theta(x_1) \ldots f_\theta(x_n) \\ &= \frac{1}{\sigma^n (\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2} \end{aligned}$$

3. A continuación, obtendremos el estadístico *razón de verosimilitudes*:

$$RV(x_1, ..., x_n) = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)}$$

Este cálculo requerirá dos cálculos previos:

Para el denominador: Obtención de la estimación de máxima verosimilitud de μ y de σ^2:

$$\hat{\mu} = \bar{x} \qquad ; \qquad \hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

Para el numerador: Obtención de la estimación de máxima verosimilitud de σ^2, para $\mu = \mu_0$ conocido:

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \mu_0)^2$$

Una vez que sabemos en qué valores se obtienen los máximos del numerador y del denominador, unos cálculos sencillos y bastante tediosos nos llevan a la siguiente expresión:

$$RV(x_1, ..., x_n) = ... = \left[\frac{1}{1 + \frac{1}{n-1} \left(\frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right)^2} \right]^{n/2}$$

Observemos que $RV(x_1, ..., x_n)$ disminuye cuando $\left| \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \right|$ aumenta. Esto lo usaremos en el siguiente paso.

4. En consecuencia, la región, R, de rechazo de la hipótesis nula, será de la forma:

$$R = \left\{ RV(x_1, ..., x_n) = \left[\frac{1}{1 + \frac{1}{n-1} \left(\frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right)^2} \right]^{n/2} \text{ es “pequeño”} \right\}$$

$$= \left\{ \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| \text{ es “grande”} \right\} = \left\{ \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| > c \right\}$$

Para determinar el valor de c, recordamos que $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$ (según vimos en los intervalos de confianza), e imponemos que el nivel de significación queremos que sea α:

$$\alpha = \max_{\theta \in \Theta_0} P_\theta(R) = \max_{\theta \in \Theta_0} P_\theta \left\{ \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| > c \right\} = P\left\{ |t_{n-1}| > c \right\}$$

Por tanto:

$$c = t_{n-1,\alpha/2}$$

En resumen:

$$R = \left\{ (x_1, ..., x_n) : \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| > t_{n-1,\alpha/2} \right\}$$

$$= \left\{ (x_1, ..., x_n) : |\bar{x} - \mu_0| > t_{n-1,\alpha/2} \ \frac{s}{\sqrt{n}} \right\}$$

Es muy interesante poner de manifiesto un par de propiedades del contraste de hipótesis que acabamos de obtener:

1. Este contraste de hipótesis se plantea para decidir si es razonable aceptar que $\mu = \mu_0$. Dado que la estimación de máxima verosimilitud de μ, en este caso, es $\hat{\mu} = \bar{x}$, es intuitivamente razonable que rechacemos el valor $\mu = \mu_0$, cuando $\bar{x}$ se aleje “demasiado” de μ_0.

   La ventaja de la metodología de los contrastes de hipótesis es que se cuantifica automáticamente que significa el que $\bar{x}$ se aleje “demasiado” de μ_0.

2. El test que acabamos de obtener está estrechamente ligado al correspondiente intervalo de confianza para estimar μ (al nivel $1 - \alpha$), cuando σ es desconocido. Observemos, para ver esto, que la región de aceptación de H_0, en el test que acabamos de obtener es, simplemente:

$$A = \left\{ (x_1, ..., x_n) : |\bar{x} - \mu_0| < t_{n-1,\alpha/2} \ \frac{s}{\sqrt{n}} \right\}$$

$$= \left\{ (x_1, ..., x_n) : -t_{n-1,\alpha/2} \ \frac{s}{\sqrt{n}} < \bar{x} - \mu_0 < t_{n-1,\alpha/2} \ \frac{s}{\sqrt{n}} \right\}$$

$$= \left\{ (x_1, \ldots, x_n) : \bar{x} - t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} < \mu_0 < \bar{x} + t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \right\}$$

   En consecuencia, una forma alternativa y equivalente de aplicar el test de razón de verosimilitudes para contrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$, al nivel de significación α, sería:

   Aceptamos $H_0 : \mu = \mu_0$, si

$$\mu_0 \in IC_{1-\alpha}(\mu) = \left(\bar{x} - t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \ ; \bar{x} + t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \right)$$

Rechazamos $H_0 : \mu = \mu_0$, si

$$\mu_0 \notin IC_{1-\alpha}(\mu) = \left(\bar{x} - t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \; ; \bar{x} + t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \right)$$

Observaciones generales sobre los contrastes de hipótesis

1. Aplicando de manera sistemática este método de razón de verosimilitudes, iríamos obteniendo los tests de hipótesis que se utilizan en las situaciones más habituales:

   Una muestra aleatoria de una característica con distribución Normal, Bernoulli, Poisson,...

   Dos muestras aleatorias independientes de características con distribución Normal, Bernoulli,...

   La mayoría de los libros dedicados a la Estadística Aplicada incluyen un listado de los contrastes de hipótesis más frecuentemente utilizados.

2. En general, cualquier contraste de hipótesis del tipo $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$, al nivel de significación α, se puede abordar de dos formas alternativas y equivalentes:

   - Se aplica el método de razón de verosimilitudes, se obtiene la región R, y decidimos de la siguiente forma:
     Si $(x_1, ..., x_n) \in R \quad \Rightarrow \quad$ Rechazamos H_0
     Si $(x_1, ..., x_n) \notin R \quad \Rightarrow \quad$ Aceptamos H_0
   - Obtenemos el intervalo de confianza para estimar θ, al nivel de confianza $1 - \alpha$, y decidimos de la siguiente forma:
     Si $\theta_0 \in IC_{1-\alpha}(\theta) \quad \Rightarrow \quad$ Aceptamos H_0
     Si $\theta_0 \notin IC_{1-\alpha}(\theta) \quad \Rightarrow \quad$ Rechazamos H_0

4 Elección de hipótesis nula y alternativa

Las secciones anteriores han estado dedicadas a como resolver el problema de contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$. Ahora bien, en un problema real, nadie nos va a decir cómo elegir las hipótesis nula y alternativa. Esto es algo que debe asignar el usuario, en función de la pregunta a la que tenga que responder. Vamos a dar un par de pautas, ya que las situaciones que se pueden presentar se pueden dividir en dos grandes grupos:

1. En unos casos, la pregunta que se nos plantea es del siguiente tipo:

   ¿Resulta aceptable que $\theta = \theta_0$?

   En este caso, tenemos que elegir entre las dos siguientes posiblilidades: $\theta = \theta_0$ y $\theta \neq \theta_0$. La asignación de papeles, en un caso de estos, es siempre de la siguiente forma:

   $H_0 : \theta = \theta_0 \qquad ; \qquad H_1 : \theta \neq \theta_0 \quad ,$

   asignando el "igual a" a la hipótesis nula.

2. En otros casos, sin embargo, la pregunta que se nos plantea es del siguiente tipo:

   ¿Disponemos de suficiente evidencia estadística para concluir que $\theta > \theta_0$?

   En este caso, tenemos que elegir entre las dos siguientes posiblilidades: $\theta > \theta_0$ y $\theta < \theta_0$. La asignación de papeles, en un caso de estos, es de la siguiente forma:

   $H_0 : \theta \leq \theta_0 \qquad ; \qquad H_1 : \theta > \theta_0 \quad ,$

   asignando lo que se está estudiando si se puede concluir a la hipótesis alternativa. La razón para proceder así es que el método de razón de verosimilitudes es muy conservador con la hipótesis nula, de modo que, si conseguimos rechazar la hipótesis nula y aceptar la hipótesis alternativa (donde hemos situado lo que se está estudiando si se puede concluir), es porque los datos apoyan muy fuertemente a la hipótesis alternativa.

   Por supuesto, si la pregunta planteada fuera del tipo:

   ¿Disponemos de suficiente evidencia estadística para concluir que $\theta < \theta_0$?

   la asignación de papeles sería de la forma:

   $H_0 : \theta \geq \theta_0 \qquad ; \qquad H_1 : \theta < \theta_0 \quad ,$

   utilizando el mismo tipo de razonamiento.

5 P-valor

El *p-valor* es una herramienta muy valiosa, automáticamente proporcionada por el ordenador, cuando se realiza un contraste de hipótesis mediante un paquete estadístico. El *p-valor* de una muestra para contrastar H_0 frente a H_1 se puede definir como "la probabilidad (evaluada bajo H_0) que tendríamos de obtener un resultado menos favorable a la hipótesis nula que el proporcionado por la muestra obtenida". Sin la ayuda del ordenador, el cálculo del p-valor suele ser pesado. Por este motivo, lo que más nos interesa de este concepto es su interpretación, una vez que nos ha sido proporcionado por el ordenador. La interpretación más sencilla y manejable es la siguiente:

El p-valor de una muestra es el apoyo que los datos proporcionan a la hipótesis nula.

Para decidir si este apoyo es suficiente o no, lo comparamos con el nivel de significación:

1. Si el p-valor está por debajo del nivel de significación elegido, el apoyo a H_0 es escaso, y la hipótesis nula debe rechazarse.

2. Si, por el contrario, el p-valor está por encima del nivel de significación elegido, el apoyo a H_0 es suficiente, y la hipótesis nula no puede rechazarse.

El p-valor y la región de rechazo, R, de un test son herramientas equivalentes, en el sentido de que la decisión alcanzada es la misma con cualquiera de ellas.