Tema 5

Contrastes de hipótesis estadísticas

Ejemplo: Se sospecha acerca de un posible vertido de una sustancia contaminante en un río por lo que es necesario estudiar si la concentración en el agua de dicha sustancia supera los límites admisibles. Como no es posible analizar la totalidad del agua se toman varias muestras en puntos repartidos uniformemente a lo largo del río midiendo, a continuación, la concentración de la sustancia contaminante en las muestras recogidas. A partir de los datos muestrales, ¿cómo podemos decidir si la concentración media del contaminante en el río supera los límites admisibles?

No interesa estimar la concentración media, sino conocer si dicha concentración supera los límites admisibles. Para responder a la pregunta anterior necesitaremos plantear y resolver un contraste de hipótesis adecuado.

Definición: Una hipótesis estadística es una conjetura sobre alguna característica desconocida de la población bajo estudio. En este tema será una afirmación sobre un parámetro poblacional.

Definición: se llama **hipótesis nula**, y se denota por H_0, a la hipótesis bajo estudio. La realización de un contraste sobre H_0 supone la contraposición de esta hipótesis frente a otra **hipótesis alternativa** que se denota H_1.

Definición: Una hipótesis se dice que es simple cuando se asigna un único valor al parámetro bajo estudio, no dejando ningún grado de libertad. En caso contrario, se dice que la hipótesis es compuesta.

Ejemplos: sea $X \sim N(\mu, \sigma^2)$

Hipótesis simples	Hipótesis compuestas
$\mu = 10$	$\mu \leq 10$
$\sigma^2 = 9$	$\sigma^2 \neq 9$
$\sigma = 1$	$\sigma > 1$

Definición: En un contraste paramétrico, una hipótesis compuesta unilateral es aquella de la forma $\theta < \theta_0$ o bien de la forma $\theta > \theta_0$ mientras que una hipótesis bilateral es de la forma $\theta \neq \theta_0$, siendo θ el parámetro sobre el cual se realiza el contraste y θ_0 el valor de prueba.

Generalmente se toma como H_0 una hipótesis simple y como H_1 una hipótesis compuesta (unilateral o bilateral).

Plantear un contraste de hipótesis consiste en el establecimiento de una hipótesis nula y una hipótesis alternativa. Por ejemplo, dada una variable aleatoria X que sigue una distribución $N(\mu, \sigma^2)$, podemos plantear el contraste

$$\begin{cases} H_0 : \mu = 10 \\ H_1 : \mu < 10 \end{cases}$$

La resolución de un contraste, que posteriormente veremos cómo se realiza, da como resultado tomar una de las dos decisiones siguientes:

- **Rechazar H_0:** cuando, a partir de los datos experimentales, se tiene evidencia significativa de que H_1 es cierta.
- **No rechazar H_0:** cuando, a partir de los datos experimentales, no se tiene evidencia significativa de que H_1 es cierta.

Ahora bien, como la decisión se toma a partir de una muestra podemos equivocarnos, siendo posibles cada una de las siguientes situaciones:

	H_0 cierta	H_0 falsa
No rechazar H_0	Correcto	**Error de Tipo II**
Rechazar H_0	**Error de Tipo I**	Correcto

Se denota por α a la probabilidad de cometer un *Error de Tipo I*, también llamada nivel de significación del contraste, y se denota por β a la probabilidad de cometer un *Error de Tipo II*; esto es:

$$P\left[\text{Rechazar } H_0|_{H_0 \text{ es cierta}}\right] = \alpha$$

$$P\left[\text{No rechazar } H_0|_{H_0 \text{ es falsa}}\right] = \beta$$

¿Se pueden controlar los errores?

Lo ideal sería que ambas probabilidades fueran cero o estuvieran muy próximas a cero. Aunque una probabilidad no es complementaria de la otra, esto es, en general $\alpha + \beta \neq 1$, resulta que cuando una disminuye la otra aumenta, por lo que no podemos minimizar las dos a la vez.

En la práctica lo que se suele hacer es fijar α como un valor pequeño. De este modo, la probabilidad de cometer un error de tipo I será pequeña y estará controlada. Por otra parte, aunque sería deseable que una vez fijado α el valor de β fuera lo menor posible, la consecución de este objetivo excede las pretensiones de este tema, por lo que nos limitaremos a calcular y controlar la probabilidad de error de tipo II en casos muy concretos.

Si H_1 es compuesta, la probabilidad de cometer un Error de Tipo II, β, depende del parámetro θ, esto es, $\beta = \beta(\theta)$ y representa la probabilidad de cometer un Error de Tipo II para cada valor θ de la hipótesis alternativa. Esto nos lleva a la siguiente definición.

Definición: Se define la función potencia del test como $\rho(\theta) = 1-\beta(\theta)$. Representa la probabilidad de rechazar H_0 cuando es falsa y conviene que sea lo mayor posible.

Para resolver un contraste se utiliza un estadístico T unido a una regla de decisión, que dependerá del α fijado.

La regla de decisión será de la forma: Rechazo $H_0 \Leftrightarrow$ "T verifica una condición". El recorrido de T queda dividido en dos partes:

Región de aceptación: valores de T que hacen que no se rechace H_0.

Región crítica o de rechazo: valores de T que hacen que se rechace H_0.

En lo que sigue se muestran las regiones críticas de los posibles contrastes sobre los parámetros de una población normal y de dos poblaciones normales independientes.

Contrastes de hipótesis, sobre poblaciones normales, a nivel de significación α

Contrastes sobre una población

Contrastes sobre μ con σ conocida	
Hipótesis	Región crítica
$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$	$\overline{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$ ó $\overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$
$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu > \mu_0 \end{cases}$	$\overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha}$
$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu < \mu_0 \end{cases}$	$\overline{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha}$

Contrastes sobre μ con σ desconocida	
Hipótesis	Región Crítica
$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$	$\overline{X} \leq \mu_0 - \frac{S_c}{\sqrt{n}} \cdot t_{n-1,1-\frac{\alpha}{2}}$ ó $\overline{X} \geq \mu_0 + \frac{S_c}{\sqrt{n}} \cdot t_{n-1,1-\frac{\alpha}{2}}$
$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu > \mu_0 \end{cases}$	$\overline{X} \geq \mu_0 + \frac{S_c}{\sqrt{n}} \cdot t_{n-1,1-\alpha}$
$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu < \mu_0 \end{cases}$	$\overline{X} \leq \mu_0 - \frac{S_c}{\sqrt{n}} \cdot t_{n-1,1-\alpha}$

Contrastes sobre σ^2	
Hipótesis	Región Crítica
$\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 \neq \sigma_0^2 \end{cases}$	$S_c^2 \geq \frac{\sigma_0^2 \chi^2_{n-1,1-\frac{\alpha}{2}}}{n-1}$ ó $S_c^2 \leq \frac{\sigma_0^2 \chi^2_{n-1,\frac{\alpha}{2}}}{n-1}$
$\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 > \sigma_0^2 \end{cases}$	$S_c^2 \geq \frac{\sigma_0^2 \chi^2_{n-1,1-\alpha}}{n-1}$
$\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 < \sigma_0^2 \end{cases}$	$S_c^2 \leq \frac{\sigma_0^2 \chi^2_{n-1,\alpha}}{n-1}$

Contrastes de hipótesis, sobre poblaciones normales, a nivel de significación α

Contrastes sobre dos poblaciones independientes

Contrastes sobre diferencias de medias con varianzas conocidas	
Hipótesis	Región Crítica
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y \neq \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \leq \delta_0 - \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \cdot z_{1-\frac{\alpha}{2}}$ o bien $\overline{X} - \overline{Y} \geq \delta_0 + \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \cdot z_{1-\frac{\alpha}{2}}$
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y > \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \geq \delta_0 + \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \cdot z_{1-\alpha}$
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y < \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \leq \delta_0 - \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \cdot z_{1-\alpha}$

Contrastes sobre diferencias de medias con varianzas iguales y desconocidas	
Hipótesis	Región Crítica*
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y \neq \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \leq \delta_0 - S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot t_{n_x+n_y-2,1-\frac{\alpha}{2}}$ o bien $\overline{X} - \overline{Y} \geq \delta_0 + S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot t_{n_x+n_y-2,1-\frac{\alpha}{2}}$
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y > \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \geq \delta_0 + S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot t_{n_x+n_y-2,1-\alpha}$
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y < \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \leq \delta_0 - S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot t_{n_x+n_y-2,1-\alpha}$

$^{*}S_p^2 = \frac{(n_x-1)S_{cx}^2+(n_y-1)S_{cy}^2}{n_x+n_y-2}$

Contrastes de hipótesis, sobre poblaciones normales, a nivel de significación α

Contrastes sobre dos poblaciones independientes

Contrastes sobre diferencias de medias con varianzas distintas y desconocidas	
Hipótesis	Región Crítica*
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y \neq \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \leq \delta_0 - \sqrt{\frac{S_{cx}^2}{n_x} + \frac{S_{cy}^2}{n_y}} \cdot t_{g,1-\frac{\alpha}{2}}$ o bien $\overline{X} - \overline{Y} \geq \delta_0 + \sqrt{\frac{S_{cx}^2}{n_x} + \frac{S_{cy}^2}{n_y}} \cdot t_{g,1-\frac{\alpha}{2}}$
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y > \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \geq \delta_0 + \sqrt{\frac{S_{cx}^2}{n_x} + \frac{S_{cy}^2}{n_y}} \cdot t_{g,1-\alpha}$
$\begin{cases} H_0: & \mu_x - \mu_y = \delta_0 \\ H_1: & \mu_x - \mu_y < \delta_0 \end{cases}$	$\overline{X} - \overline{Y} \leq \delta_0 - \sqrt{\frac{S_{cx}^2}{n_x} + \frac{S_{cy}^2}{n_y}} \cdot t_{g,1-\alpha}$

Contrastes sobre varianzas de dos poblaciones	
Hipótesis	Región Crítica
$\begin{cases} H_0: & \sigma_x^2 = \sigma_y^2 \\ H_1: & \sigma_x^2 \neq \sigma_y^2 \end{cases}$	$\frac{S_{cx}^2}{S_{cy}^2} \geq f_{n_x-1,n_y-1,1-\frac{\alpha}{2}}$ o bien $\frac{S_{cx}^2}{S_{cy}^2} \leq \frac{1}{f_{n_y-1,n_x-1,1-\frac{\alpha}{2}}}$
$\begin{cases} H_0: & \sigma_x^2 = \sigma_y^2 \\ H_1: & \sigma_x^2 > \sigma_y^2 \end{cases}$	$\frac{S_{cx}^2}{S_{cy}^2} \geq f_{n_x-1,n_y-1,1-\alpha}$
$\begin{cases} H_0: & \sigma_x^2 = \sigma_y^2 \\ H_1: & \sigma_x^2 < \sigma_y^2 \end{cases}$	$\frac{S_{cx}^2}{S_{cy}^2} \leq \frac{1}{f_{n_y-1,n_x-1,1-\alpha}}$

$$^{*}\, g \approx \frac{\left(\frac{S_{cx}^2}{n_x} + \frac{S_{cy}^2}{n_y}\right)^2}{\frac{\left(S_{cx}^2/n_x\right)^2}{n_x+1} + \frac{\left(S_{cy}^2/n_y\right)^2}{n_y+1}} - 2$$