Capítulo 6

Contrastes de hipótesis

1. Introducción.

En muchas ocasiones el objetivo que se persigue con la realización de un muestreo o de un experimento es poner a prueba alguna hipótesis concebida previamente. Esta es, de hecho, la esencia del método científico: observar, concebir hipótesis y contrastar dichas hipótesis con nuevas observaciones. Ahora bien si, como ocurre frecuentemente, las observaciones están expuestas a fuertes dosis de variabilidad aleatoria, resulta difícil distinguir el efecto que se desea medir de ese "ruido de fondo".

Pongamos un ejemplo sencillo: en un estudio de la morfología de cierta especie, un investigador puede tener *a priori* buenas razones para pensar que los machos deben ser, en promedio, mayores que las hembras. A partir de una muestra aleatoria de 5 machos y 5 hembras, observa en los machos un peso medio de 2,54 kg, frente a 2,77 kg de media en las hembras. ¿Contienen estos datos *evidencia suficiente* para refutar la hipótesis de partida? Es obvio que no todos los animales tienen el mismo peso –*variabilidad natural*– y que, aún siendo cierta la hipótesis de partida, cabe la posibilidad –*por efecto del azar*– de que dicha hipótesis no se verifique.

En este capítulo se desarrollarán los fundamentos básicos para la construcción de contrastes de hipótesis: métodos que, teniendo en cuenta la presencia de la variabilidad y del azar, permitan establecer reglas para decidir si, dentro de ciertos márgenes de error, los datos obtenidos por muestreo o experimentación contienen evidencia suficiente para rechazar la hipótesis de partida o si ésta puede seguir aceptándose como válida.

Una vez establecidos los fundamentos de los contrastes de hipótesis, se estudiarán en particular algunos contrastes de uso frecuente en la práctica, referidos a hipótesis sobre los parámetros de distribuciones de probabilidad conocidas.

Objetivos.

Al finalizar este capítulo el alumno deberá:

1. Conocer y comprender el concepto de contraste de hipótesis.
2. Conocer y comprender los dos tipos de error posibles en un contraste de hipótesis y por tanto los conceptos de nivel de significación y potencia.
3. Conocer, comprender y ser capaz de calcular en algunos casos el p-valor de un contraste.
4. Conocer y ser capaz de aplicar contrastes de hipótesis frecuentes en la práctica, en particular los relativos a medias, varianzas y proporciones.
5. Ser capaz de distinguir las condiciones necesarias para la aplicación de cada contraste de hipótesis.
6. Ser capaz de calcular el tamaño de muestra necesario para la realización de un contraste con significación y potencia predeterminados.
7. Ser capaz de resolver problemas prácticos de contraste de hipótesis utilizando el programa R .

2. Conceptos básicos.

En la actividad científico-técnica práctica, el objetivo que se persigue en muchas ocasiones con la realización de un muestreo o de un experimento es poner a prueba alguna hipótesis concebida previamente.

Por ejemplo:

- Se ha diseñado un nuevo método de depuración de agua, cuyas características físico-químicas inducen a suponer que reducirán la concentración de ciertos contaminantes biológicos con mayor eficiencia que el método que se venía usando hasta ahora. ¿Será verdad esta suposición?
- Se cree que cierto compuesto químico actúa sobre los peces que se crían en tanques de cultivo, reduciendo los niveles de estrés que presentan estos animales al tener que compartir un espacio reducido con un elevado número de congéneres. ¿Es cierta esta conjetura?

- Un método de análisis químico A es mucho más caro que otro método B, pero ¿es realmente mucho más preciso?
- ¿La tasa de mortalidad en cultivos marinos realizados en tanques cerrados es superior a la que se produce en cultivos en mar abierto?

Todos los ejemplos que hemos citado se caracterizan por describir situaciones en las que es imposible realizar un experimento u observación que nos confirme o desmienta *de una manera absolutamente segura* la hipótesis planteada. De ahí que los procedimientos para tomar decisiones sobre la veracidad o falsedad de estas hipótesis hayan de ser necesariamente procedimientos estadísticos, con los que se pretende mantener bajo control el riesgo de tomar decisiones erróneas.

Una hipótesis estadística es una afirmación o conjetura con respecto a alguna característica de interés de la distribución de una variable aleatoria. Llamaremos *hipótesis nula* (H_0) a la hipótesis de partida, que será aceptada como válida si la evidencia en su contra es débil o inexistente. La *hipótesis alternativa* (H_1) será la hipótesis que será aceptada en caso de que se rechace H_0.

Un *contraste de hipótesis* estadístico es una regla de decisión que permita elegir entre la dos hipótesis, H_0 y H_1, en función de la evidencia aportada por los datos disponibles y del riesgo de error que estemos dispuestos a asumir.

Las hipótesis estadísticas pueden plantearse de muy diversas formas:

- En función de los parámetros de la distribución de probabilidad. Por ejemplo, ¿el valor medio de cierta variable en una población es cero?, ¿son iguales las medias de dos poblaciones?, ¿la proporción de sujetos con cierta característica supera el 70 % de la población?
- En términos de la forma de la distribución de la variable de interés: ¿se distribuye una variable de igual forma en dos poblaciones?, ¿es normal la distribución de una variable?.
- En términos de características de asociación: ¿son dos variables independientes?, ¿la relación entre dos variables es lineal?

3. Tipos de Error en los contrastes de hipótesis.

En un contraste de hipótesis es posible cometer dos tipos de error:

Error tipo I: Rechazar la hipótesis nula cuando es verdadera

Error tipo II: Aceptar la hipótesis nula cuando es falsa.

En general, llamaremos:

$$\alpha = P\left(\text{Error Tipo I}\right) = P\left(\text{Rechazar } H_0 \left| H_0 \text{ es cierta}\right.\right)$$
$$\beta = P\left(\text{Error Tipo II}\right) = P\left(\text{Aceptar } H_0 \left| H_0 \text{ es falsa}\right.\right)$$

De esta forma, al realizar un contraste de hipótesis son posibles las siguientes situaciones:

		Realidad	
		H_0 cierta	H_0 falsa
Decisión	Aceptar H_0	Decisión correcta (1-α)	Error II (β)
	Rechazar H_0	Error I (α)	Decisión Correcta (1-β)

La probabilidad α de cometer un error tipo I se conoce como *Nivel de significación* del contraste. Asimismo, la probabilidad de no cometer un error tipo II:

$$1 - \beta = P\left(\text{Rechazar } H_0 \left| H_0 \text{ es falsa}\right.\right)$$

se conoce como *Potencia del contraste.* Ambas probabilidades, pues, miden la probabilidad de rechazar la hipótesis nula: α cuando es cierta y $1-\beta$ cuando es falsa. La situación ideal es que α sea lo más pequeña posible y $1-\beta$ lo más grande posible. Ello en la práctica se traduce en tener mucha información (muchos datos). Cuando no es posible disponer de toda la información que sería deseable (situación muy frecuente en los estudios reales) en general se procurará que α sea pequeña, aún a costa de que β pueda ser grande (y por ende $1-\beta$ pequeña).

4. Contrastes de Significación.

Supongamos que se desea decidir si el valor (desconocido) de cierto parámetro θ pertenece o no a un conjunto Θ_0. Este parámetro está asociado a la distribución de probabilidad de cierta variable aleatoria X, de la que es posible extraer una muestra aleatoria $(X_1, X_2, \ldots, X_n)$ que contiene información sobre θ. El procedimiento general de los contrastes o pruebas de significación es el siguiente:

1. Fijar las hipótesis nula (H_0 :$\theta \in \Theta_0$) y alternativa (H_1 : $\theta \notin \Theta_0$).
2. Determinar un *estadístico de contraste* dependiente de los datos, $T\left(X_1, X_2, \ldots, X_n\right)$, cuya distribución de probabilidad sea conocida cuando H_0 es cierta.
3. Fijar la probabilidad α de error de tipo I (*nivel de significación del* contraste), y determinar una *región crítica* R_C de tal manera que:

$$P\left(T\left(X_1, X_2, \ldots, X_n\right) \in R_C \left| H_o \text{ es cierta}\right.\right) = \alpha$$

4. Obtener una muestra aleatoria $\left(X_1, X_2, \ldots, X_n\right)$ y utilizar la siguiente regla de decisión:

Si $T\left(X_1, X_2, \ldots, X_n\right) \in R_C$ *rechazar* H_0*. En caso contrario aceptar* H_0*.*

Observaciones:

1. Con esta regla de decisión se tiene que la probabilidad de error tipo I es:

$$\begin{aligned} P\left(\text{Error Tipo I}\right) &= P\left(\text{Rechazar } H_0 \left| H_0 \text{ es cierta}\right.\right) = \\ &= P\left(T\left(X_1, X_2, \ldots, X_n\right) \in R_C \left| H_o \text{ es cierta}\right.\right) = \alpha \end{aligned}$$

2. Al mismo tiempo, la probabilidad de error tipo II queda, en principio, indeterminada:

$$\begin{aligned} P\left(\text{Error Tipo II}\right) &= P\left(\text{Aceptar } H_0 \left| H_0 \text{ es falsa}\right.\right) = \\ &= P\left(T\left(X_1, X_2, \ldots, X_n\right) \notin R_C \left| H_o \text{ es falsa}\right.\right) \end{aligned}$$

   si bien, como veremos, puede calcularse para las alternativas de interés, e incluso prefijarse de antemano, fijando un tamaño de muestra adecuado.

3. Para entender el fundamento de los contrastes de significación tengamos en cuenta que, una vez tomados los datos, sólo pueden ocurrir dos cosas: que T caiga en R_C o que no lo haga. Entonces:

a) Si $T \notin R_C$ estaría ocurriendo algo que era muy probable que ocurriese si H_0 fuera cierta ya que, tal como se ha definido R_C, se tiene que:

$$P\left(T\left(X_1, X_2, \ldots, X_n\right) \notin R_C \left| H_o \text{ es cierta}\right.\right) = 1 - \alpha$$

Por tanto, el resultado del test en este caso es el esperado si H_0 es cierta, por lo que nada se opone a aceptar dicha hipótesis. Nótese, no obstante, que aceptar H_0 *no significa* que hayamos demostrado que H_0 sea cierta, sino sólo que los datos no la contradicen. Dicho de otra forma *aceptamos H_0 no porque hayamos podido probar que es cierta, sino porque no hemos podido probar que es falsa.*

b) Si $T \in R_C$ estaría ocurriendo algo que, de ser H_0 cierta, muy difícilmente podía haber ocurrido. Pero como de hecho ha ocurrido, ello nos indica que los datos contienen una fuerte evidencia de que H_0 es posiblemente falsa o, lo que es lo mismo, una fuerte evidencia de que H_1 es posiblemente cierta.

4. Nótese la no simetría de las dos posibles conclusiones del contraste:

   a) Cuando se acepta H_0 es porque la evidencia en su contra es débil.

   b) Cuando se acepta H_1 es porque la evidencia a su favor es fuerte.

   Por esta razón, cuando planteamos un contraste de hipótesis se debe colocar como hipótesis alternativa aquella de la que queramos tener fuerte evidencia a su favor en caso de que finalmente sea aceptada. La hipótesis nula, en cambio, es la que se aceptará por defecto si no hay fuerte evidencia en su contra (e incluso si no hay fuerte evidencia a su favor).

   Por todo ello, cuando un test concluye con la aceptación de H_0 se dice que ha resultado *no significativo,* y cuando concluye con su rechazo se dice que ha resultado *significativo.*

5. La región crítica R_C suele denominarse también *región de rechazo* (de H_0). La región complementaria se denomina *Región de Aceptación, R_A.* Obviamente

$$P\left(T\left(X_1, X_2, \ldots, X_n\right) \in R_A \left| H_o \text{ es cierta}\right.\right) = 1 - \alpha$$

   La región de aceptación contiene, pues, los valores del estadístico $T\left(X_1, X_2, \ldots, X_n\right)$ que, con mucha probabilidad, podrían observarse *por puro azar* si H_0 fuese cierta.

Ejemplo 6.1.

Las algas de cierta especie que se cultivan con fines farmacológicos son muy sensibles al pH del agua. Se ha observado que el desarrollo de estas algas es óptimo cuando el pH promedio es 1, y diariamente se realizan controles con el objetivo de aplicar medidas correctoras (añadir aditivos químicos al agua) si el pH se aparta de este valor. Estos controles consisten en tomar 5 muestras de agua y evaluar el pH medio. En un día en que el pH medio de las cinco muestras es de 1.2 con una desviación típica de 0.4. ¿sería preciso aplicar alguna medida correctora? (se supone que la distribución del pH es normal)

1. Si llamamos μ al pH medio real del agua, el problema puede plantearse como el contraste de hipótesis:
$$\begin{cases} H_0 : \mu = 1 \\ H_1 : \mu \neq 1 \end{cases}$$
siendo la información disponible la aportada por una muestra de cinco valores de pH, $\{X_1, X_2, X_3, X_4, X_5\}$.

2. Como no conocemos el valor de μ, podemos estimarlo mediante la media muestral $\bar{X}$. Si H_0 fuera verdad, entonces el valor de $\bar{X}$ debería parecerse a 1. Ello significa que la hipótesis nula H_0 debería rechazarse si $\bar{X}$ se aleja de 1, esto es, si $\left|\bar{X} - 1\right|$ es un valor grande. ¿Como de grande? Para responder a esta pregunta observemos que si H_0 es cierta se tiene que:
$$T\left(X_1, \ldots, X_5\right) = \frac{\bar{X} - 1}{s/\sqrt{5}} \approx t_4$$

3. Podemos usar ahora la tabla de la t de Student para encontrar el valor $t_{4,\alpha/2}$ tal que:
$$P\left(\left|\frac{\bar{X} - 1}{s/\sqrt{5}}\right| > t_{4,\alpha/2} \middle| H_0 \text{ cierta}\right) = \alpha$$
De esta forma, la región crítica es $R_C = \left(-\infty, -t_{4,\alpha/2}\right] \cup \left[t_{4,\alpha/2}, \infty\right)$.

4. El contraste consiste entonces en *rechazar* H_0 *si* $\frac{\bar{X}-1}{s/\sqrt{5}} \in R_C$ *y aceptar* H_0 *en caso contrario.* Con los datos de este ejemplo se obtiene $\frac{\bar{X}-1}{s/\sqrt{5}} = \frac{1{,}2-1}{0{,}4/\sqrt{5}} = 1{,}11$. Asimismo, si elegimos $\alpha = 0{,}05$ resulta $t_{4,0{,}025} = 2{,}776$. Como el valor 1.11 no está en la región de rechazo concluimos que puede aceptarse H_0.

Dicho de otra forma, *si* H_0 *fuera cierta*, sería muy improbable que $\left|\frac{\bar{X}-1}{s/\sqrt{5}}\right| > 2{,}776$; o de manera equivalente, lo mas probable sería que $\left|\frac{\bar{X}-1}{s/\sqrt{5}}\right| \leq 2{,}776$. Como el valor observado, 1.11, está dentro de lo que es muy probable observar cuando H_0 es cierta, concluimos que no existe evidencia suficiente para rechazar H_0.

Ejemplo 6.2. Supongamos ahora que las algas de nuestro ejemplo se desarrollan bien si $\mu \leq 1$, pero mueren si $\mu > 1$, siendo μ el pH medio del agua del tanque de cultivo. Si en 7 análisis de agua hemos obtenido un pH medio de 1.1, con desviación típica 0.3, ¿hay evidencia suficiente para rechazar H_0?

En este caso, el contraste que se plantea es de la forma:

$$\begin{cases} H_0: & \mu \leq 1 \\ H_1: & \mu > 1 \end{cases}$$

Obviamente, aún siendo cierta H_0 podría ocurrir por azar que la media muestral $\bar{X}$ fuese *algo mayor* que 1, pero no *mucho mayor.* Por tanto la hipótesis nula H_0 debería rechazarse si el valor de $\bar{X} - 1$ es más grande de lo que cabría esperar por azar cuando $\mu \leq 1$. Para determinar como de grande debe ser $\bar{X} - 1$ para rechazar H_0 podemos utilizar como estadístico de contraste:

$$T(X_1, \ldots, X_7) = \frac{\bar{X} - 1}{S/\sqrt{7}}$$

Cuando H_0 es cierta, el valor de μ para el que cabría esperar valores más altos de $\bar{X}$ por azar es $\mu = 1$, en cuyo caso el estadístico $T(X_1, \ldots, X_7)$ sigue una distribución t de Student con 6 grados de libertad. Por tanto tenemos que:

$$P\left(\frac{\bar{X} - 1}{S/\sqrt{7}} > t_{6,\alpha} \middle| \mu = 1\right) = \alpha$$

Además, si $\mu < 1$ esta probabilidad será más pequeña y por tanto:

$$P\left(\frac{\bar{X} - 1}{S/\sqrt{7}} > t_{6,\alpha} \middle| H_0 \text{ cierta}\right) = P\left(\frac{\bar{X} - 1}{S/\sqrt{7}} > t_{6,\alpha} \middle| \mu \leq 1\right) \leq \alpha$$

De esta forma, si H_0 es cierta, es muy difícil que $T(X_1, \ldots, X_7)$ sea mayor que $t_{6,\alpha}$, por lo que la región crítica o de rechazo para este test es $R_C = [t_{6,\alpha}, \infty)$. Si $T(X_1, \ldots, X_7)$ cayera en este intervalo estaría ocurriendo algo muy difícil de ser H_0 cierta, por lo que H_0 debe rechazarse.

Con los datos aportados en el ejemplo se obtiene $\frac{\bar{X}-1}{S/\sqrt{7}} = \frac{1{,}1-1}{0{,}3/\sqrt{7}} = 0{,}882$. Asimismo, si elegimos $\alpha = 0{,}05$ resulta $t_{6,0{,}05} = 1{,}943$ y la región crítica es $R_C = [1{,}943, \infty)$. Como el valor 0.882 no está en esta región concluimos que puede aceptarse H_0.

Nota: Los contrastes de la forma $\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases}$ reciben el nombre de *contrastes bilaterales o de dos colas* (su región crítica es bilateral). Los contrastes de la forma $\begin{cases} H_0: & \theta \geq \theta_0 \\ H_1: & \theta < \theta_0 \end{cases}$ ó $\begin{cases} H_0: & \theta \leq \theta_0 \\ H_1: & \theta > \theta_0 \end{cases}$ se denominan *contrastes unilaterales o de una cola.*

4.1. P-valor de un contraste.

Tal como hemos visto, en la construcción del contraste de hipótesis juega un papel importante el *nivel de significación* α. Este valor representa la probabilidad *que consideramos aceptable* de cometer un error tipo I: rechazar la hipótesis nula cuando es cierta. En este sentido, el valor de α es arbitrario. En el ámbito científico es habitual utilizar los valores 0.05, 0.01 e incluso 0.001. Pero cualquier otro valor podría ser igualmente válido (en la práctica hay consenso en que, en cualquier caso, α nunca debe ser mayor que 0.1).

Obviamente, cuanto más pequeño sea el valor de α, más difícil es rechazar H_0 cuando es cierta. Una vez obtenida una muestra $X_1, X_2, \ldots X_n$, se define el *p-valor* del contraste como el valor mínimo de α para el cual es posible rechazar H_0 con esos datos. Así, por ejemplo:

- Si con los datos disponibles, el valor más pequeño de α que permite el rechazo de H_0 es 0.4, ello querría decir que sólo sería posible rechazar H_0 si estuviéramos dispuestos a aceptar una probabilidad del 40 % de rechazarla siendo cierta (lo que obviamente no resultaría razonable).

- Si con los datos disponibles, el valor mínimo de α que conduce al rechazo de H_0 es 0.02, ello significa que sería posible rechazar esta hipótesis incluso si exigimos un riesgo del 2 % de rechazarla siendo cierta; pero no podríamos rechazarla si el riesgo asumible fuese del 1 %.

De esta forma, una vez obtenida la muestra, podríamos basar nuestra decisión en la siguiente regla basada en el p-valor:

Si $p - valor \geq \alpha$ aceptar H_0. Si $p - valor < \alpha$ rechazar H_0

Ejemplo 6.3. La región crítica para el rechazo de H_0 en el ejemplo 6.1 era de la forma $R_C = \left(-\infty, -t_{4,\alpha/2}\right] \cup \left[t_{4,\alpha/2}, \infty\right)$. Con los datos del ejemplo, el valor del estadístico de contraste fue $\frac{\bar{X}-1}{s/\sqrt{5}} = 1{,}11$. El valor más pequeño de α que permitiría entonces el rechazo de H_0 sería el que produjese $t_{4,\alpha/2} = 1{,}11$ (para que la región de rechazo contenga al valor del estadístico de contraste). Para hallar este valor de α basta tener en cuenta que, por definición:

$$P\left(t_4 \geq t_{4,\alpha/2}\right) = \frac{\alpha}{2}$$

Por tanto

$$P\left(t_4 \geq 1{,}11\right) = \frac{\alpha}{2}$$

La tabla de la t de Student no permite calcular esta probabilidad de forma sencilla, pero podemos calcularla con R :

$$P\left(t_4 \geq 1{,}11\right) = 1 - P\left(t_4 < 1{,}11\right) = \texttt{1-pt(1.11,4)} = 0{,}1646$$

Así pues:

$$\frac{\alpha}{2} = 0{,}1646$$

de donde:

$$\alpha = 0{,}3292$$

De esta forma, para los datos del ejemplo, el *p-valor* (valor mínimo de α que conduce al rechazo de H_0) es 0.3292. Siguiendo la regla del p-valor, sólo rechazaríamos H_0 si estuviésemos dispuestos a asumir una probabilidad 0.3292 de rechazar dicha hipótesis siendo cierta. Como no es el caso (habíamos elegido $\alpha = 0{,}05$), aceptamos H_0.

5. Potencia de un contraste.

Tal como hemos señalado, cuando se realiza un contraste de significación, la regla de decisión se establece de tal forma que el riesgo de cometer un error tipo I –rechazar la hipótesis nula cuando es cierta– es como mucho α, el nivel de significación del test. De esta forma, si se rechaza la hipótesis nula, sabemos *a priori* que existe muy poco riesgo de equivocarnos. Pero ¿qué ocurre si se acepta la hipótesis nula? ¿cuál es el riesgo de aceptar una hipótesis nula falsa? La probabilidad de cometer este error (error tipo II) es la que hemos denotado como β. Su valor complementario $1-\beta$ recibe el nombre de *potencia del contraste* y representa la probabilidad de rechazar H_0 cuando es falsa. Tal como hemos definido los contrastes de significación:

$$1-\beta = P\left(T\left(X_1, X_2, \ldots, X_n\right) \in R_C \left| H_o \text{ es falsa}\right.\right)$$

Ejemplo 6.4. Con los datos del ejemplo 6.1 en el contraste:

$$\begin{cases} H_0 : \mu = 1 \\ H_1 : \mu \neq 1 \end{cases}$$

hemos aceptado la hipótesis nula ($\mu = 1$) aún cuando la media muestral era 1.2. ¿Cuál es la probabilidad de que estemos cometiendo un error de tipo II en este contraste? Para responder a esta pregunta observemos que esta probabilidad es:

$$\begin{aligned} P\left(\text{Error Tipo II}\right) &= P\left(\text{Aceptar } H_0 \left| H_0 \text{ es falsa}\right.\right) = \\ &= P\left(T\left(X_1, X_2, \ldots, X_n\right) \notin R_C \left| H_o \text{ es falsa}\right.\right) = \\ &= P\left(\left|\frac{\bar{X}-1}{s/\sqrt{5}}\right| \leq t_{4,\alpha/2} \Big/ \mu \neq 1\right) = P\left(-t_{4,\alpha/2} \leq \frac{\bar{X}-1}{s/\sqrt{5}} \leq t_{4,\alpha/2} \Big/ \mu \neq 1\right) \end{aligned}$$

Para calcular esta probabilidad hemos de tener en cuenta que realizamos el contraste bajo el supuesto de que la variable X que se mide (en este caso el pH) es $N\left(\mu, \sigma\right)$, por lo que el estadístico

$$\frac{\bar{X}-\mu}{s/\sqrt{5}}$$

sigue una distribución t de Student con 4 grados de libertad. Cuando H_0 es falsa se tiene que $\mu \neq 1$ y por tanto:

$$\begin{aligned} \beta\left(\mu\right) =& P\left(-t_{4,\alpha/2} \leq \frac{\bar{X}-1}{s/\sqrt{5}} \leq t_{4,\alpha/2} \Big/ \mu \neq 1\right) = \\ & P\left(-t_{4,\alpha/2} \leq \frac{\bar{X}-\mu+\mu-1}{s/\sqrt{5}} \leq t_{4,\alpha/2} \Big/ \mu \neq 1\right) = \\ & P\left(-t_{4,\alpha/2} \leq \frac{\bar{X}-\mu}{s/\sqrt{5}} + \frac{\mu-1}{s/\sqrt{5}} \leq t_{4,\alpha/2} \Big/ \mu \neq 1\right) = \\ & P\left(-t_{4,\alpha/2} - \frac{\mu-1}{s/\sqrt{5}} \leq \frac{\bar{X}-\mu}{s/\sqrt{5}} \leq t_{4,\alpha/2} - \frac{\mu-1}{s/\sqrt{5}} \Big/ \mu \neq 1\right) = \\ & P\left(-t_{4,\alpha/2} - \frac{\mu-1}{s/\sqrt{5}} \leq t_4 \leq t_{4,\alpha/2} - \frac{\mu-1}{s/\sqrt{5}} \Big/ \mu \neq 1\right) \end{aligned}$$

Así pues, la probabilidad de error tipo II corresponde, geométricamente, al área bajo la función de densidad de una t_4 entre los valores $-t_{4,\alpha/2} - \frac{\mu-1}{s/\sqrt{5}}$ y $t_{4,\alpha/2} - \frac{\mu-1}{s/\sqrt{5}}$. La figura 1 muestra gráficamente esta área para diversos valores de μ.

Tal como puede apreciarse en esta figura, a medida que el valor de μ se aleja de 1, el término $\frac{\mu-1}{s/\sqrt{5}}$ se hace mayor en valor absoluto, por lo que el intervalo $\left[-t_{4,\alpha/2}-\frac{\mu-1}{s/\sqrt{5}},\, t_{4,\alpha/2}-\frac{\mu-1}{s/\sqrt{5}}\right]$ se va desplazando (hacia la izquierda si $\mu>1$, o hacia la derecha si $\mu<1$). Como consecuencia de este desplazamiento, el área que comprende la función de densidad sobre este intervalo –esto es, el valor de la probabilidad de error II, β– se va haciendo cada vez menor. La interpretación de este comportamiento de β es bastante intuitiva: en nuestro contraste estamos tratando de decidir si la verdadera media de la población es 1; será más fácil equivocarse aceptando que es 1 cuando realmente es 0.9 ó 1.1 (el verdadero valor μ está cerca de 1) que cuando la verdadera media es un valor más alejado de 1, como el 0.2 ó el 1.8.

Podemos también calcular numéricamente los valores de β para diversos valores alternativos de μ. Para el contraste del ejemplo 6.1 habíamos elegido $\alpha = 0{,}05$, resultando $t_{4,0{,}025} = 2{,}776$; asimismo, teníamos que $s = 0{,}4$. Por tanto, la probabilidad de error tipo II en este caso es, dependiendo del valor de μ:

$$\begin{aligned}\beta(\mu) =& P\left(-2{,}776-\frac{\mu-1}{0{,}4/\sqrt{5}} \leq t_4 \leq 2{,}776-\frac{\mu-1}{0{,}4/\sqrt{5}}\right) = \\ & P\left(t_4 \leq 2{,}776-\frac{\mu-1}{0{,}4/\sqrt{5}}\right) - P\left(t_4 \leq -2{,}776-\frac{\mu-1}{0{,}4/\sqrt{5}}\right)\end{aligned}$$

La tabla de la t de Student no se presta a calcular estas probabilidades, pero podemos utilizar R :

$\beta(\mu)$=`pt(2.776-(mu-1)/(0.4/sqrt(5)),4)-pt(2.776-(mu-1)/(0.4/sqrt(5)),4)`

La siguiente tabla muestra los valores de la probabilidad de error tipo II, así como la potencia que se alcanza para diversos valores de μ:

μ	$\beta(\mu)$	Potencia = $1-\beta(\mu)$
0	0.0235	0.9765
0.2	0.0816	0.9184
0.4	0.2953	0.7047
0.6	0.6873	0.3127
0.8	0.9049	0.0951
1	0.95	0.05
1.2	0.9049	0.0951
1.4	0.6873	0.3127
1.6	0.2953	0.7047
1.8	0.0816	0.9184
2.0	0.0235	0.9765

Asimismo, la figura 2 representa gráficamente estos valores, mostrando las funciones de error tipo II y potencia para este contraste. En esta figura vemos nuevamente que la probabilidad de error tipo II, $\beta(\mu)$, es tanto mayor cuanto más próximo esté μ a 1, alcanzando su máximo cuando μ coincide con el valor especificado en la hipótesis nula ($\mu = 1$). El comportamiento de la función de potencia –probabilidad de rechazar H_0 cuando es falsa– es, como cabe esperar, justo en inverso: si el verdadero valor de μ está cerca de 1, el contraste apenas tiene potencia para distinguir ambos valores; cuánto más lejos esté μ de 1, mayor es la potencia del contraste.

6. Tamaño de muestra para una significación y potencia preespecificadas.

El contraste de hipótesis que hemos planteado en el 6.1 es un caso particular de contraste de la forma:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

en el que la hipótesis nula que se pone a prueba es si puede aceptarse que el valor de la esperanza μ de una variable X con distribución normal es μ_0. Si se dispone de una muestra aleatoria de n observaciones de esta variable, siendo $\bar{X}$ su media y S su desviación típica, la regla de decisión para este contraste, fijado un nivel de significación α es, generalizando el procedimiento que hemos visto en el ejemplo 6.1:

Rechazar H_0 si $\left|\frac{\bar{X}-\mu_0}{s/\sqrt{n}}\right| > t_{n-1,\alpha/2}$ y aceptar H_0 en caso contrario.

Asimismo, generalizando el resultado obtenido en el ejemplo 6.4, la probabilidad de error tipo II para este contraste viene dada por:

$$\beta(\mu) = P\left(-t_{n-1,\alpha/2} - \frac{\mu-\mu_0}{s/\sqrt{n}} \le t_{n-1} \le t_{n-1,\alpha/2} - \frac{\mu-\mu_0}{s/\sqrt{n}} \Big/ \mu \neq \mu_0\right) \tag{6.1}$$

que, como ya hemos visto, representa el área comprendida por la densidad t de Student con $n-1$ grados de libertad sobre el intervalo $\left[-t_{n-1,\alpha/2} - \frac{\mu-\mu_0}{s/\sqrt{n}},\, t_{n-1,\alpha/2} - \frac{\mu-\mu_0}{s/\sqrt{n}}\right]$ (ver figura 1). Obsérvese que este intervalo puede expresarse también de la forma:

$$\left[-t_{n-1,\alpha/2} - \frac{(\mu-\mu_0)\sqrt{n}}{s},\, t_{n-1,\alpha/2} - \frac{(\mu-\mu_0)\sqrt{n}}{s}\right]$$

lo que hace evidente el hecho de que aún cuando $\frac{(\mu-\mu_0)}{s}$ tomase un valor pequeño, eligiendo un valor adecuado de n (tamaño de la muestra) podemos hacer el término $\frac{(\mu-\mu_0)\sqrt{n}}{s}$ todo lo grande que queramos. Ello significa que, tal como vimos en nuestro análisis de la figura 1, podemos desplazar el intervalo anterior (hacia la izquierda o la derecha, según el signo de $\mu-\mu_0$) hasta que el área comprendida sobre el mismo –esto es, la probabilidad de error II– sea tan pequeña como se quiera.

Esto nos permite responder a la cuestión siguiente: *¿cuál debe ser el tamaño n de la muestra si se desea que cuando $\mu=\mu_0+\Delta$ la probabilidad de error tipo II en el contraste anterior sea un valor prefijado β –o, de modo equivalente, que la potencia sea 1-β–, manteniendo al mismo tiempo un nivel de significación preespecificado α?*

Para ello, utilizando la ecuación 6.1, y teniendo en cuenta que $\mu-\mu_0=\Delta$, debemos encontrar el valor de n tal que:

$$\begin{aligned}
\beta =& P\left(-t_{n-1,\alpha/2} - \frac{\Delta\sqrt{n}}{s} \le t_{n-1} \le t_{n-1,\alpha/2} - \frac{\Delta\sqrt{n}}{s}\right) = \\
=& P\left(t_{n-1} > -t_{n-1,\alpha/2} - \frac{\Delta\sqrt{n}}{s}\right) - P\left(t_{n-1} > t_{n-1,\alpha/2} - \frac{\Delta\sqrt{n}}{s}\right) \cong \\
\cong& P\left(Z > -z_{\alpha/2} - \frac{\Delta\sqrt{n}}{s}\right) - P\left(Z > z_{\alpha/2} - \frac{\Delta\sqrt{n}}{s}\right) \cong \\
\cong& 1 - P\left(Z > z_{\alpha/2} - \frac{\Delta\sqrt{n}}{s}\right) \Rightarrow P\left(Z > z_{\alpha/2} - \frac{\Delta\sqrt{n}}{s}\right) = 1-\beta
\end{aligned}$$

(aquí hemos hecho dos aproximaciones; en primer lugar hemos supuesto que n va a resultar tan grande que la distribución t_n puede aproximarse por la normal estándar Z; y en segundo lugar hemos supuesto que el valor $-z_{\alpha/2} - \frac{\Delta\sqrt{n}}{S}$ es tan grande en valor absoluto que el área a su derecha es prácticamente uno). Utilizando la notación habitual z_β para el percentil de la normal estándar tal que $P(Z > z_{1-\beta}) = 1 - \beta$ tenemos que:

$$z_{\alpha/2} - \frac{\Delta\sqrt{n}}{s} = z_{1-\beta} = -z_\beta$$

de donde, despejando n, resulta:

$$n = \frac{\left(z_{\alpha/2} + z_\beta\right)^2 s^2}{\Delta^2}$$

Obsérvese que el valor de n:

- Es proporcional al cuadrado de la suma $z_{\alpha/2} + z_\beta$. Como estos valores son más grandes a medida que α y β son más pequeños, el tamaño de la muestra se incrementa cuando se desea que las probabilidades de los errores I y II disminuyan.

- Es proporcional a la varianza s^2, por lo que cuanto mayor sea la variabilidad en la variable que se mide mayor habrá de ser el tamaño de la muestra. Es intuitivamente claro que debe ser así. Si los valores de X fuesen muy homogéneos (poca variabilidad), una muestra pequeña podría representar bien el comportamiento de la variable; a medida que los valores de X son más heterogéneos será precisa más información –más datos– para representarla.

- Es inversamente proporcional al cuadrado de la diferencia Δ que se pretende detectar entre el verdadero valor medio μ y el valor μ_0 que se pone a prueba. Ello significa que cuanto menor sea la diferencia que se pretende detectar, mayor habrá de ser el tamaño de muestra.

El valor de s^2 no se conoce habitualmente antes de realizar el muestreo, por lo que para planificar el tamaño adecuado de muestra, habrá que utilizar un valor de s^2 obtenido en una muestra piloto o publicado en la literatura en estudios similares.

Señalemos por último que en esta sección hemos desarrollado el cálculo del tamaño de la muestra sólo para contrastar si el valor esperado μ de una variable es igual a un valor preespecificado μ_0. No obstante, el mismo patrón de ideas se aplica para el cálculo del tamaño muestral en otros contrastes de hipótesis, con las lógicas modificaciones derivadas del tipo de datos y de la forma de la regla de decisión. Asimismo, las observaciones que se acaban de realizar sobre la relación del tamaño de muestra con las magnitudes de α, β, Δ y la variabilidad resultan de aplicación general en todos los contrastes de hipótesis.

Ejemplo 6.5. Volviendo al ejemplo 6.1, recordemos que el crecimiento de las algas allí descritas requiere que el pH medio del agua sea 1. Supongamos además que las algas tienen cierta tolerancia a variaciones en el pH y que su desarrollo en cualquier caso es óptimo si el pH medio se mantiene entre 0.8 y 1.2. Se desea planificar el número de muestras de agua diarias que deben tomarse si se desea realizar el contraste

$$\begin{cases} H_0 : \mu = 1 \\ H_1 : \mu \neq 1 \end{cases}$$

con un nivel de significación 0.05, y garantizando una potencia 0.9 de que se rechazará H_0 si μ cae por debajo de 0.8 o por encima de 1.2.

Usando la información aportada por la muestra del ejemplo 6.1, usaremos como estimador piloto de la varianza el valor $s^2 = 0{,}4^2 = 0{,}16$. La diferencia mínima que interesa detectar en este caso es $\Delta = 0{,}2$, ya que se nos dice que las algas muestran tolerancia con valores de pH que difieran de 1 en 0.2 unidades (entre 0.8 y 1.2). Dado que se desea detectar esta diferencia con potencia $1 - \beta = 0{,}9$, se tiene $\beta = 0{,}1$ y $z_\beta = z_{0,1} = 1{,}28$. Para el nivel de significación $\alpha = 0{,}05$ se tiene $z_{\alpha/2} = 1{,}96$, y por tanto:

$$n = \frac{\left(z_{\alpha/2} + z_\beta\right)^2 s^2}{\Delta^2} = \frac{(1{,}96 + 1{,}28)^2 \cdot 0{,}16}{0{,}2^2} \cong 42$$

7. Significación estadística y relevancia práctica.

Ya hemos señalado más arriba que cuando en un contraste se rechaza la hipótesis nula, tal resultado se suele expresar diciendo que *el contraste ha resultado significativo.* Es necesario tener aquí cierta precaución con la terminología, ya que la palabra *"significativo"* en este contexto suele ser mal interpretada. La definición que proporciona el diccionario del adjetivo "*significativo" es* "*que tiene importancia por representar o significar algo".* Por ello, el hablante habitual cuando emplea esta palabra la entiende normalmente como referida a algo *importante*. Sin embargo, en el contexto de un contraste de hipótesis estadístico, el que un resultado haya sido significativo indica simplemente *que dicho resultado no puede explicarse como efecto del azar.* Que sea importante o no, es algo que habrá de ser valorado en función de las implicaciones prácticas que pueda tener dicho resultado.

Así, en el ejemplo 6.5 hemos visto que con una muestra de 42 observaciones del pH del agua hay una probabilidad del 90 % de detectar si el pH medio difiere en más de 0.2 unidades del valor medio deseado $\mu = 1$. El lector puede utilizar la misma fórmula para comprobar que, con la misma potencia, si la muestra fuese de tamaño 672 se podría detectar una diferencia de 0.05 unidades, y con una muestra de 1867 observaciones se podría detectar una diferencia de 0.03 unidades. Ello significa que, si se hacen los correspondientes contrastes de hipótesis con esos tamaños muestrales, las diferencias citadas, en caso de encontrarse, serían declaradas "*significativas*". Pero desde luego no serían *importantes*: si las algas se desarrollan bien cuando el pH medio se aparta hasta 0.2 unidades de 1, ¿qué importancia tendría haber encontrado que el pH medio es *significativamente* distinto de 1 porque se aparta de ese valor en 0.03 unidades?

Así pues, en general con una muestra lo suficientemente grande cualquier diferencia puede resultar estadísticamente significativa, por muy irrelevante que su valor resulte en la práctica. Obviamente también es cierto lo contrario: si la muestra es demasiado pequeña, diferencias importantes pueden resultar no significativas (recuérdese: aceptar la hipótesis nula no significa que sea cierta). Es responsabilidad del investigador, por tanto, fijar la diferencia mínima Δ que se considera relevante o importante y determinar el tamaño de muestra para que se pueda detectar dicha diferencia con una significación y potencia adecuados. Sólo en estas condiciones podrá ser el resultado de un contraste significativo y relevante a la vez.

8. Relación entre intervalos de confianza y contrastes de hipótesis.

En el capítulo anterior hemos estudiado la construcción de intervalos de confianza para los parámetros de ciertas distribuciones de probabilidad. Recordemos que $[\theta_1(\mathfrak{X}), \theta_2(\mathfrak{X})]$, donde $\theta_1(\mathfrak{X})$ y $\theta_2(\mathfrak{X})$ son variables aleatorias que dependen de una muestra $\mathfrak{X} = \{X_1, X_2, \ldots, X_n\}$, es un *intervalo de confianza a nivel* $1-\alpha$ *para el parámetro* θ si la probabilidad de que el intervalo contenga a dicho parámetro es $1-\alpha$, esto es:

$$P\left(\theta \in [\theta_1(\mathfrak{X}), \theta_2(\mathfrak{X})]\right) = 1-\alpha$$

Entonces, si se dispone de un intervalo de confianza para θ, para resolver el contraste de hipótesis:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases}$$

puede utilizarse como regla de decisión:

Si $\theta_0 \in [\theta_1(\mathfrak{X}), \theta_2(\mathfrak{X})]$ *aceptar* H_0*; en caso contrario, rechazar* H_0.

En efecto, la probabilidad de error tipo I cuando se utiliza esta regla es:

$$\begin{aligned} P\left(\text{error I}\right) &= P\left(\text{rechazar } H_0 \left| H_0 \text{ cierta}\right.\right) = \left(\theta_0 \notin \left[\theta_1\left(\mathfrak{X}\right), \theta_2\left(\mathfrak{X}\right)\right] \left|\theta = \theta_0\right.\right) = \\ &= P\left(\theta \notin \left[\theta_1\left(\mathfrak{X}\right), \theta_2\left(\mathfrak{X}\right)\right]\right) = \alpha \end{aligned}$$

Ejemplo 6.6. En el ejemplo 6.1 debíamos decidir, a partir de 5 muestras de pH de un tanque de agua, si podía aceptarse que el pH medio era 1. Para ello planteábamos el contraste:

$$\begin{cases} H_0: & \mu = 1 \\ H_1: & \mu \neq 1 \end{cases}$$

partiendo del supuesto adicional de que el pH sigue una distribución normal. El intervalo de confianza para la media μ de una distribución normal con varianza σ^2 desconocida es, tal como vimos en el capítulo anterior:

$$\left[\bar{X} - \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}\right]$$

Por tanto, podríamos utilizar como regla de decisión para el contraste:

Si $1 \in \left[\bar{X} - \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}\right]$*, aceptar* H_0 *y en caso contrario rechazar* H_0*.*

Es fácil comprobar que:

$$1 \in \left[\bar{X} - \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}\right] \Leftrightarrow \bar{X} - \frac{s}{\sqrt{n}} t_{n-1,\alpha/2} \leq 1 \leq \bar{X} + \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}$$
$$\bar{X} - 1 - \frac{s}{\sqrt{n}} t_{n-1,\alpha/2} \leq 0 \leq \bar{X} - 1 + \frac{s}{\sqrt{n}} t_{n-1,\alpha/2} \Leftrightarrow -\frac{s}{\sqrt{n}} t_{n-1,\alpha/2} \leq \bar{X} - 1 \leq \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}$$
$$-t_{n-1,\alpha/2} \leq \frac{\bar{X} - 1}{s/\sqrt{n}} \leq t_{n-1,\alpha/2} \Leftrightarrow \left|\frac{\bar{X} - 1}{s/\sqrt{n}}\right| \leq t_{n-1,\alpha/2}$$

Por tanto la regla de decisión basada en el intervalo de confianza *es exactamente la misma* que ya habíamos obtenido en el ejemplo 6.1 por otro procedimiento.

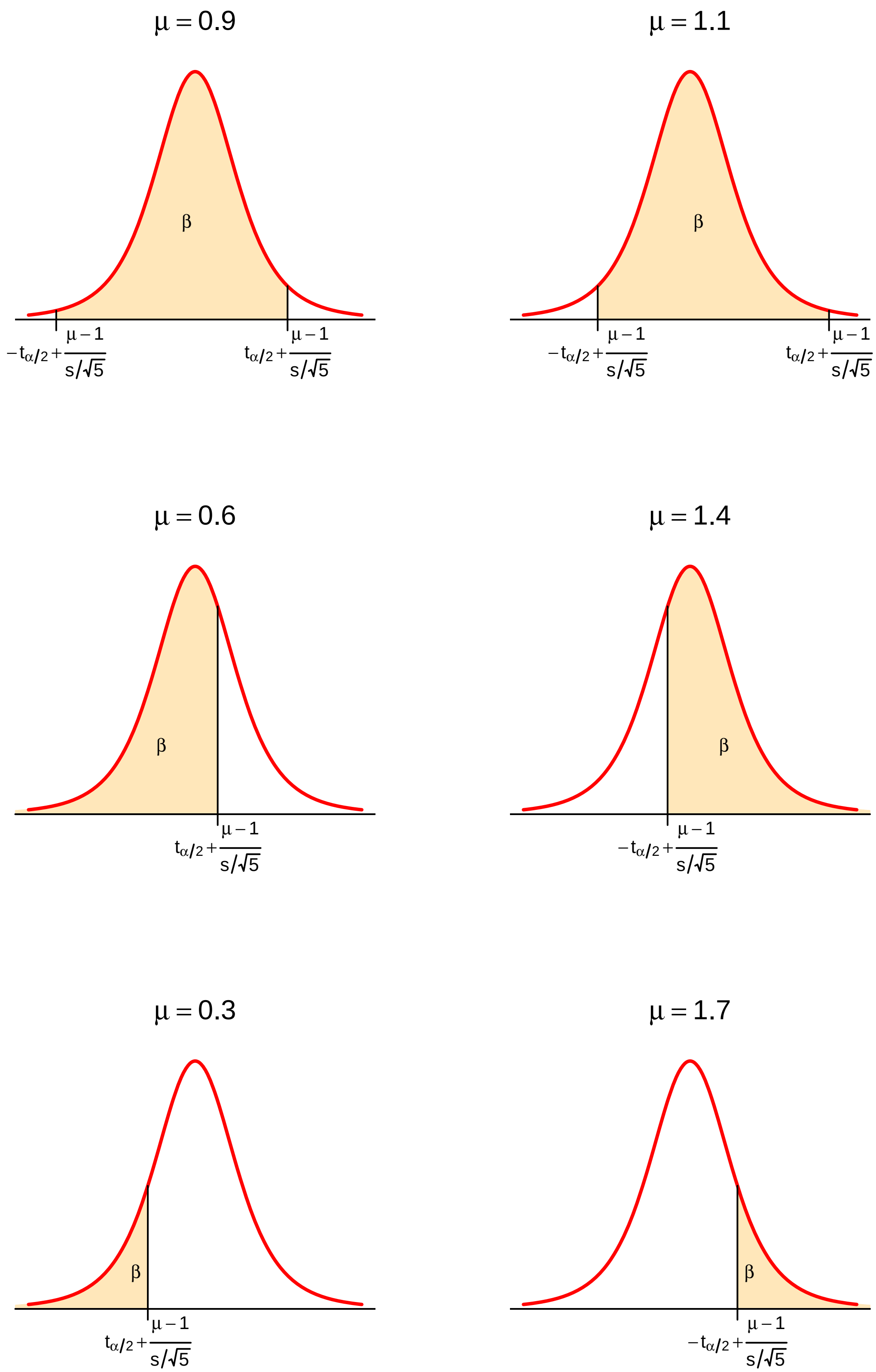

Figura 1: Probabilidad de error tipo II para diversos valores de μ en el contraste de hipótesis del ejemplo 6.1.

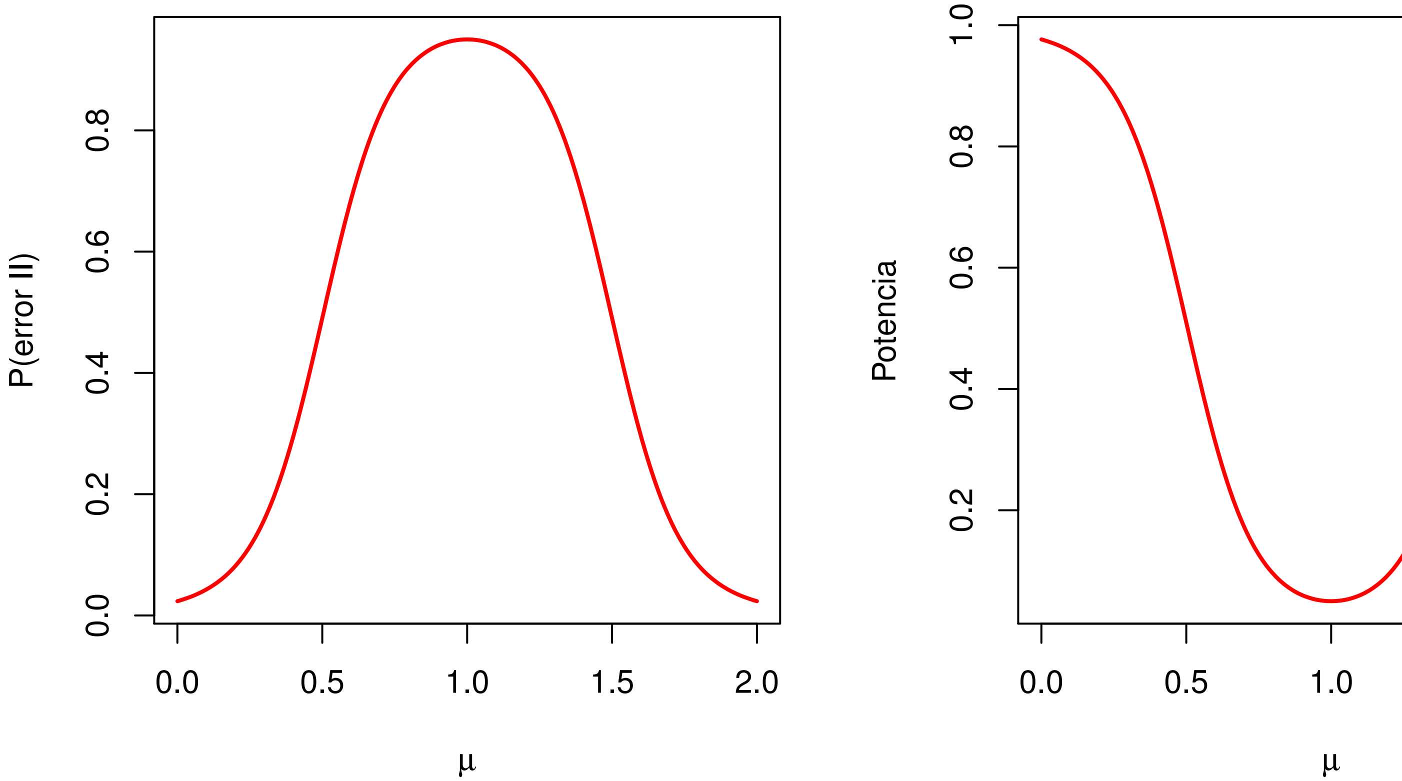

Figura 2: Funciones de error tipo II (izquierda) y potencia (derecha) para el contraste de hipótesis del ejemplo 6.1