

FACULTAD DE CIENCIAS
GRADO EN MATEMÁTICAS
TRABAJO FIN DE GRADO
CURSO ACADÉMICO 2023-2024

TÍTULO:

Métodos clásicos y bayesianos para determinar el tamaño muestral en investigación clínica

AUTOR:

Alejandro Sellés Pérez

Resumen

Este trabajo realiza un recorrido por diversas técnicas aplicadas en investigación clínica en un momento previo a la realización de un estudio, con el objetivo de la obtener el tamaño muestral necesario para realizar un buen análisis. Se tratan varios métodos basados en distintas corrientes estadísticas.

En la primera parte se estudian los criterios clásicos basados en la estadística frecuentista. Por un lado, se analizan los métodos para realizar la estimación de una media o una diferencia de medias. Y, por otro lado, se deducen las fórmulas para estudios cuyo fin es realizar una comparación de medias aplicando un análisis de potencia.

La segunda parte del trabajo se centra en estudiar tres criterios basados en la estadística bayesiana para la estimación de un parámetro. Además, se realiza una comparación entre estos métodos y se estudian las diferencias que existen entre ellos.

La sección final se dedica a realizar una comparativa entre todas las fórmulas obtenidas para los ensayos clínicos cuyo fin es obtener la estimación de una media. El objetivo es analizar si es interesante realizar este tipo de análisis previos al estudio desde un punto de vista bayesiano.

Palabras clave: Tamaño muestral, Investigación clínica, Contrastes de Hipótesis, Métodos Bayesianos.

Abstract

This project reviews various techniques applied in clinical research prior to carrying out a study, with the aim of obtaining the sample size necessary to carry out a good analysis. Several methods based on different statistical trends are discussed.

In the first part, the classical criteria based on frequentist statistics are studied. On the one hand, the methods for estimating a mean or a difference of means are analysed. On the other hand, the formulas for studies whose purpose is to make a comparison of means are deduced by applying a power analysis.

The second part of the study focuses on three criteria based on Bayesian statistics for the estimation of a parameter. In addition, a comparison is made between these methods and the differences between them are studied.

The final section is devoted to a comparison of all the formulae obtained for clinical trials aimed at obtaining the estimation of a mean. The aim is to analyse whether it is interesting to perform this type of pre-study analysis from a Bayesian point of view.

Key words: Sample Size, Clinical Research, Hypothesis Testing, Bayesian Methods.

Índice

1. Introducción	6
2. Preliminares	8
2.1. Teoría frecuentista	8
2.2. Teoría bayesiana	11
3. Métodos frecuentistas	15
3.1. Estimación de medias	15
3.1.1. Una muestra	15
3.1.2. Dos muestras	17
3.1.3. Ejemplo	19
3.2. Compación de medias	19
3.2.1. Diseño con una muestra	21
3.2.1.1. Test de Igualdad	22
3.2.1.2. Test de no inferioridad o superioridad	24
3.2.2. Diseño paralelo con dos muestras	26
3.2.2.1. Test de Igualdad	27
3.2.2.2. Test de no inferioridad o superioridad	28
3.2.3. Ejemplo	30
4. Métodos bayesianos	31
4.1. Motivación	31
4.2. 3 criterios de selección	32
4.2.1. Average Coverage Criterion (ACC)	32
4.2.2. Average Length Criterion (ALC)	33
4.2.3. Worst Outcome Criterion (WOC)	33
4.3. Estimación de una media. Criterios y comparación	34
4.3.1. Precisión conocida	34
4.3.2. Precisión desconocida	35

4.3.3. Comparación de criterios	38
4.4. Estimación de una diferencia de medias. Criterios y comparación	43
4.4.1. Precisión conocida	43
4.4.2. Precisión desconocida	44
4.4.3. Comparación de criterios	46
5. Comparación entre métodos bayesianos y frecuentistas	48
6. Conclusiones	50
Referencias	52
A. Detalles del desarrollo del trabajo	54
B. Código de R	55

1. Introducción

La investigación clínica desempeña un papel fundamental en el avance de la medicina, proporcionando evidencia sólida para la toma de decisiones en el ámbito de la salud. Uno de los aspectos cruciales en el diseño de estudios clínicos es la determinación del tamaño muestral adecuado, debido a que este no solo afecta directamente a la fiabilidad de los resultados obtenidos o a los costes económicos, sino que también puede afectar negativamente a la salud de los individuos que participan en la investigación. Podemos ver como ejemplo un ensayo clínico del año 2002, donde se realizó un estudio de un nuevo fármaco, llamado AN-1792, cuyo fin era tratar el Alzheimer. Tras un tiempo, se vieron obligados a interrumpir la investigación en EEUU y varios países europeos, entre ellos España, después de que 4 de los 360 participantes en un fase II desarrollasen signos de inflamación en su sistema nervioso central.

Por otro lado, existen algunos estudios donde el tamaño de la muestra que podemos obtener puede ser muy pequeño; por ejemplo los relacionados con enfermedades raras que padecen muy pocos individuos. En estos ensayos también es importante saber qué conclusiones se pueden obtener con este número reducido de personas.

A lo largo de este trabajo abordaremos el tema de la obtención de un tamaño muestral óptimo desde dos enfoques principales, a menudo enfrentados: la estadística frecuentista y la estadística bayesiana. El objetivo principal será explorar y comparar estas teorías. Para ello, se revisarán los fundamentos teóricos de los métodos, así como sus ventajas, limitaciones y consideraciones prácticas.

Por un lado, los métodos clásicos frecuentistas han sido ampliamente utilizados en la investigación clínica durante décadas y se basan en supuestos estadísticos bien establecidos. Estas fórmulas suelen requerir la especificación de parámetros como el tamaño del efecto, la variabilidad de la población y el nivel de confianza deseado. Sin embargo, veremos cómo presentan ciertas limitaciones y cómo éstas podrían ser abordadas aplicando otro tipo de técnicas. Para la obtención de las fórmulas, aplicaremos el método de la cantidad pivotal, el test de razón de verosimilitudes o el teorema de Karlin-Rubin.

Por otro lado, la estadística bayesiana ha ganado popularidad en los últimos años gracias al desarrollo computacional y a la gran cantidad de datos almacenados con los que contamos actualmente, ya que estos pueden ser incluidos como información previa a la hora de realizar los estudios. Es decir, en lugar de depender únicamente de supuestos poblacionales, los métodos bayesianos permiten a los investigadores integrar el conocimiento experto y los resultados de estudios piloto en la determinación del tamaño muestral. Esta

aproximación ofrece una mayor flexibilidad y puede conducir a estimaciones más precisas, especialmente en situaciones donde los datos son escasos o heterogéneos.

Nos centraremos en poblaciones normales y en diseños de experimentos con una muestra o diseños paralelos con dos muestras, esto es, experimentos que, por un lado cuentan con un grupo placebo o de control, al que le suministran el medicamento ya establecido o un tratamiento que no tiene ningún efecto; y, por otro lado, cuentan con un grupo de test, en el que se encontrarán los individuos sometidos a las nuevas técnicas a estudiar. También, a lo largo del trabajo se proporcionarán ejemplos ilustrativos y se discutirán casos de aplicación en diferentes contextos clínicos.

Cabe destacar que en investigación clínica se suele escoger un tamaño muestral entre un 5 % y un 10 % mayor al obtenido por los métodos, debido a que a la hora de realizar un experimento con personas, siempre hay que tener en cuenta posibles bajas a lo largo del tiempo. Además, en el momento de aplicar las fórmulas que vamos a obtener, si el valor del tamaño muestral que recibimos es un valor no entero, tomaremos el menor entero mayor al número obtenido.

En resumen, este estudio pretende ofrecer una visión integral y actualizada sobre los métodos frecuentistas y bayesianos disponibles a la hora de determinar el tamaño muestral en algunas clases de experimentos en investigación clínica, con el fin de ayudar a los investigadores a tomar decisiones correctas de manera previa al comienzo de un ensayo.

2. Preliminares

A lo largo de esta sección definiremos los conceptos, teoremas y métodos que vamos a utilizar para obtener las fórmulas. Por un lado, para la parte frecuentista, vamos a tomar como referencia los apuntes de la asignatura de Inferencia Estadística y el libro [12]. Y, por otro lado, en lo referente a la parte bayesiana, utilizaremos principalmente como referencia [7].

2.1. Teoría frecuentista

Vamos a analizar distintos métodos para la obtención del tamaño muestral, es decir, el número de sujetos que componen la muestra extraída de una población. Desde el punto de vista de la estadística frecuentista, vamos a tomar experimentos cuyo fin es la estimación de un parámetro y experimentos cuyo fin es la comparación de parámetros.

Intervalos de confianza

Por un lado, para realizar la estimación de un parámetro θ , la teoría frecuentista propone la obtención de un intervalo de confianza, para el cual definiremos previamente un valor $\alpha \in]0, 1[$, donde $1 - \alpha$ será el nivel de confianza. Esto significa, por ejemplo, tomando $\alpha = 0,05$, pensar que 95 de cada 100 veces el parámetro real estará dentro del intervalo obtenido.

Para obtener este intervalo de confianza aplicaremos el método de la cantidad pivotal. En primer lugar, buscaremos un pivote $T(X_1, \dots, X_n, \theta)$ para el parámetro de interés θ , es decir, una variable aleatoria cuya distribución no dependa de él. Y, posteriormente calcularemos dos constantes K_1, K_2 tales que

$$P_{\theta}(K_1 < T(X_1, \dots, X_n, \theta) < K_2) = 1 - \alpha.$$

Después, despejamos θ hasta llegar a

$$P_{\theta}(T_1(X_1, \dots, X_n) < \theta < T_2(X_1, \dots, X_n)) = 1 - \alpha.$$

Y así, ya tenemos $]T_1(x_1, \dots, x_n), T_2(x_1, \dots, x_n)[$ un intervalo de confianza para θ con nivel de confianza $1 - \alpha$.

Finalmente, para nuestro estudio será importante tener en cuenta que el intervalo de confianza varía su longitud l en función del valor de α escogido; esta longitud es conocida

también como la *precisión del intervalo*. Naturalmente, intervalos largos proporcionarán estimaciones imprecisas, mientras que intervalos cortos proporcionan estimaciones más precisas.

Contrastes de hipótesis

Por otro lado, para realizar una comparación entre dos parámetros, los autores frecuentistas proponen resolver un contraste de hipótesis, el cual consiste en una partición del espacio paramétrico $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$, planteada de la forma:

$$\left. \begin{array}{l} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{array} \right\},$$

donde H_0 es la hipótesis nula y H_1 es la hipótesis alternativa. La manera de resolverlo es encontrar una regla para decidir si aceptar o rechazar la hipótesis nula. Es decir, dados (x_1, \dots, x_n) , tomaremos un test $\varphi(x_1, \dots, x_n)$, que representará la probabilidad de rechazar H_0 . A la hora de trabajar con ello, debemos tener en cuenta los posibles errores que podemos cometer cuando tomamos una decisión.

En primer lugar, puede ocurrir que aceptemos la hipótesis nula H_0 cuando realmente es falsa, esto será un error de tipo II, y la probabilidad de cometerlo será denotada con β .

En segundo lugar, podemos rechazar H_0 cuando realmente es cierta, esto será un error de tipo I y denotaremos con α la probabilidad de cometerlo.

En la Tabla 1 podemos observar un resumen con las decisiones y errores que podemos cometer a la hora de realizar un contraste de hipótesis.

	Aceptar H_0	Rechazar H_0
Verdadero H_0	Decisión correcta	Error tipo I
Verdadero H_1	Error tipo II	Decisión correcta

Tabla 1: Tipos de errores en pruebas de hipótesis.

Lo ideal sería minimizar simultáneamente las probabilidades de cometer ambos errores. Sin embargo, esto no es posible, ya que al reducir la probabilidad de cometer un tipo de error, aumenta la probabilidad de cometer el otro. Por ello, en la práctica, dependiendo del estudio que estemos realizando, vamos a centrarnos en intentar disminuir un tipo de error u otro. En los tipos de ensayos de investigación clínica en los que vamos a centrarnos, cuando realicemos un contraste de hipótesis con el fin de comparar medias, intentaremos mantener un nivel razonable de α y buscaremos minimizar la probabilidad de cometer el

error tipo II, esto se conoce como análisis de potencia. Habitualmente, los investigadores toman los valores de $\alpha = 0,05$ y $\beta = 0,20$.

A la hora de encontrar la región crítica del test aplicaremos distintos métodos. En primer lugar, para resolver los test de igualdad de la forma

$$\left. \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{array} \right\} ,$$

aplicaremos el test de razón de verosimilitudes. Para ello, debemos obtener la razón de verosimilitudes:

$$\lambda(x_1, \dots, x_n) = \frac{\sup_{\Theta_0} L_{(x_1, \dots, x_n)}(\theta)}{\sup_{\Theta} L_{(x_1, \dots, x_n)}(\theta)} = \frac{L_{(x_1, \dots, x_n)}(\hat{\theta}_0)}{L_{(x_1, \dots, x_n)}(\hat{\theta})} ,$$

donde $L_{(x_1, \dots, x_n)}(\theta) = f_{\theta}(x_1, \dots, x_n)$ es la función de verosimilitud y $\hat{\theta}$ es el estimador de máxima verosimilitud de θ .

Una vez obtenida esta función, el test de razón de verosimilitudes propone tomar la región crítica, es decir, la región de rechazo de la hipótesis nula H_0 :

$$\chi_1 = \{(x_1, \dots, x_n) | \lambda(x_1, \dots, x_n) < k\} ,$$

donde $k \in (0, 1)$ y es calculada de manera que el tamaño del test sea menor o igual que α . Es decir, $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$, con $\beta(\theta) = E_{\theta}(\varphi)$.

En segundo lugar, para resolver los test de no inferioridad o superioridad de la forma

$$\left. \begin{array}{l} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{array} \right\} ,$$

buscaremos un test uniforme de máxima potencia (TUMP). Para ello, en primer lugar, debemos analizar si se cumple la razón de verosimilitudes respecto del estadístico $T(x_1, \dots, x_n)$; es decir, si dados $\theta_1 < \theta_2$,

$$\frac{f_{\theta_2}(x_1, \dots, x_n)}{f_{\theta_1}(x_1, \dots, x_n)} ,$$

tomando $+\infty$ si el denominador es 0, es una función monótona creciente respecto de T .

Ahora, como vamos a trabajar siempre con poblaciones normales, pertenecerán a la familia exponencial uniparamétrica, esto quiere decir que existirán las funciones $a(\theta) > 0$, $b(x_1, \dots, x_n) > 0$, $q(\theta) \neq 0$ y $T(x_1, \dots, x_n)$ tales que

$$f_{\theta} = a(\theta)b(x_1, \dots, x_n)e^{q(\theta)T(x_1, \dots, x_n)}, \quad \forall \theta \in \Theta, \quad \forall (x_1, \dots, x_n)$$

Por ello, para encontrar el TUMP, bastará simplemente con aplicar los siguientes resultados:

Proposición 2.1 *Si X pertenece a la familia exponencial uniparamétrica y $q(\theta)$ es estrictamente creciente, entonces se cumple la RVM respecto de T .*

Teorema 2.2 (Karlin-Rubin) *Supongamos que se cumple la RVM respecto del estadístico T . Entonces un TUMP con nivel de significación α para el contraste de hipótesis:*

$$\left. \begin{array}{l} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{array} \right\},$$

tiene una función crítica

$$\varphi^*(x_1, \dots, x_n) = \begin{cases} 1, & \text{si } T(x_1, \dots, x_n) > K \\ \gamma, & \text{si } T(x_1, \dots, x_n) = K \\ 0, & \text{si } T(x_1, \dots, x_n) < K \end{cases},$$

donde el valor de K se calcula de manera que $\beta^(\theta_0) = \alpha$.*

2.2. Teoría bayesiana

A diferencia de la estadística frecuentista donde el parámetro de interés a estudiar es una constante desconocida, desde la visión bayesiana consideraremos los parámetros como variables aleatorias, y partiremos en nuestro estudio de información sobre estos parámetros. Esta información suele venir de estudios previos realizados en los que el parámetro en cuestión ha sido estudiado, o directamente de la opinión subjetiva de expertos en la materia; y estará resumida en una distribución de probabilidad, que será la llamada *distribución a priori*.

La idea de la teoría bayesiana es, partir de esta distribución a priori sobre un parámetro θ , y, aplicando el Teorema de Bayes, obtener una distribución posterior que contendrá información sobre el parámetro actualizada, con la que se terminará trabajando.

En primer lugar, vamos a ver entonces lo que nos dice este teorema sobre el que se sustenta toda la teoría bayesiana.

Teorema 2.3 (Teorema de Bayes) *Sean B_1, B_2, \dots, B_n eventos mutuamente excluyentes y exhaustivos, con $P(B_i) \neq 0$, $i = 1, 2, \dots, n$. Para cualquier evento nuevo A , tal que $P(A) \neq 0$, tenemos:*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^n P(A|B_i)P(B_i)} \propto P(A|B_i)P(B_i).$$

Teorema 2.4 (*Teorema de Bayes*) Sean x y θ variables aleatorias con funciones de densidad $f(x|\theta)$ y $f(\theta)$. Tenemos

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_{\Theta} f(x|\theta)f(\theta)d\theta} \propto f(x|\theta)f(\theta) ,$$

donde

- x : vector o matriz de datos.
- θ : parámetro desconocido.
- $f(x_1, \dots, x_n|\theta)$: verosimilitud de los datos dado θ .
- $f(\theta)$: distribución a priori del parámetro θ .

Esta distribución obtenida a partir del Teorema 2.4, $f(\theta|x)$, es la llamada *distribución a posteriori*.

Observación 2.5 *La idea principal del mantra bayesiano es que la densidad a posteriori es proporcional a la densidad a priori por la verosimilitud. Es decir,*

$$f(\theta|x_1, \dots, x_n) \propto f(x_1, \dots, x_n|\theta)f(\theta) .$$

Observación 2.6 *Aplicando el Teorema de Bayes obtenemos una actualización de $f(\theta)$ a $f(\theta|x_1, \dots, x_n)$.*

A lo largo de nuestro estudio, vamos a trabajar suponiendo que las muestras siguen una distribución normal. Y, vamos a diferenciar los casos en los que el valor de la varianza σ^2 sea o no conocida. Por ello, con el fin de facilitar nuestros posteriores cálculos, vamos a denominar con el nombre de *precisión* a τ , donde

$$\tau = \frac{1}{\sigma^2} .$$

Este es un concepto ampliamente utilizado por los autores bayesianos, y no debe ser confundido con la precisión del intervalo de confianza vista anteriormente.

Además, vamos a definir el concepto de distribuciones conjugadas, ya que también simplificará los cálculos a la hora de obtener la distribución a posteriori, debido a que previamente conoceremos la familia de funciones a la que pertenecerá la distribución que estamos buscando.

Definición 2.7 *Supongamos que una distribución a priori es de cierta clase o familia de distribuciones paramétricas. Diremos que esa distribución es conjugada respecto de la verosimilitud si la distribución a posteriori también pertenece a esa clase de familias paramétricas.*

Concretamente, a lo largo del trabajo vamos a trabajar siempre con distribuciones normales, por lo que las distribuciones a posteriori que obtendremos seguirán siendo normales.

Además, en los casos en los que supongamos que la precisión τ es desconocida, nos veremos forzados a obtener la distribución marginal de la media μ . Para ello, partiremos de una densidad a priori para el parámetro conjunto (μ, τ) ; y aplicando el Teorema de Bayes obtendremos la distribución a posteriori conjunta $f(\mu, \tau|x)$. Y, una vez hemos obtenido esta distribución, buscaremos la distribución marginal de la media μ resolviendo:

$$f(\mu|x) = \int_0^{+\infty} f(\mu, \tau|x) d\tau$$

Tras lograr obtener la distribución a posteriori, nuestra idea será obtener el tamaño muestral necesario para realizar una buena estimación de un parámetro. Como hemos visto anteriormente, en la teoría frecuentista esto se realiza a partir de intervalos de confianza. Sin embargo, en la teoría bayesiana encontramos algunas diferencias, debido a que no hay estimadores y no existen los intervalos de confianza. En estadística bayesiana, dado un parámetro desconocido θ , se conoce como *intervalo creíble al $(1 - \alpha)100\%$* o *intervalo de credibilidad posterior al $(1 - \alpha)100\%$* , al intervalo cuya probabilidad de contener al parámetro θ sea de $1 - \alpha$. Este intervalo es obtenido a partir de la probabilidad posterior, y se suele llamar también intervalo de credibilidad posterior.

Observación 2.8 *Los intervalos creíbles son análogos a los intervalos de confianza de la estadística frecuentista, aunque difieren en una base filosófica, pues los intervalos bayesianos tratan sus límites como fijos y el parámetro estimado como una variable aleatoria, mientras que los intervalos de confianza frecuentista tratan sus límites como variables aleatorias y el parámetro como un valor fijo.*

Se pretende que el intervalo sea lo más preciso posible, entonces se tomará como intervalo de credibilidad aquel de probabilidad $1 - \alpha$ que tenga menor amplitud, y por tanto, que tenga mayor densidad de probabilidad. Este es el llamado *intervalo de mayor densidad de probabilidad posterior (HDPI)*. Concretamente, a lo largo del trabajo, como vamos a trabajar con distribuciones normales, este intervalo tendrá una densidad única y simétrica.

Para terminar, hemos recopilado las principales razones que expresan diversos autores defendiendo la idea de realizar estudios desde una visión bayesiana. Algunos de los razonamientos más comunes son:

- En primer lugar, la estadística bayesiana es una manera muy natural de implementar el método científico. Partimos de una distribución a priori que representa nuestras creencias iniciales acerca del modelo, y obtenemos la distribución posterior que representa las creencias actualizadas una vez tenemos los datos.
- En segundo lugar, a la hora de realizar estimaciones, los intervalos de confianza suelen ser a menudo malinterpretados, ya que se tiende a confundir el nivel de confianza con la probabilidad de encontrar el parámetro en el intervalo. Sin embargo, el análisis desde una perspectiva bayesiana es más fácil de comprender, ya que aquí sí hablaremos de la probabilidad de que un parámetro se encuentre dentro del intervalo calculado.
- En tercer lugar, a la hora de la práctica, los estudios frecuentistas muchas veces no evitan totalmente la subjetividad. Por ejemplo, la forma en la que el problema está planteado o el modo de recolectar los datos son elecciones subjetivas que influyen a la hora de hacer el estudio. Concretamente, esto se dará a la hora de obtener el tamaño muestral, debido a que este estudio se realiza de manera previa a la obtención de los datos. Por ello, el punto de vista bayesiano es muy interesante, ya que mucha de esta información a priori estará expresada de manera explícita y se integrará en el proceso del análisis.

Finalmente, en la Tabla 2 podemos ver una recopilación de las principales diferencias entre ambas teorías estadísticas.

	Teoría Frecuentista	Teoría Bayesiana
<i>Parámetro de interés</i>	Es una constante desconocida	Es una variable aleatoria
<i>Distribución a priori</i>	No se tiene en cuenta	Existe y es necesaria para los cálculos
<i>Distribución a posteriori</i>	No existe	Existe y se deriva
<i>Tipo de razonamiento</i>	Inductivo	Deductivo

Tabla 2: Diferencias entre teorías.

3. Métodos frecuentistas

3.1. Estimación de medias

A lo largo de esta sección vamos a buscar una fórmula que, dados unos datos que establezcamos previamente a nuestro estudio, nos devuelva el número necesario de personas para realizar una correcta estimación de un parámetro. Concretamente, este parámetro a estudiar será la media μ o la diferencia de dos medias $\mu_1 - \mu_2$.

Para alcanzar estos tamaños muestrales, los autores deterministas proponen obtener un intervalo de confianza para el parámetro en cuestión; y, posteriormente, despejar el valor de n en función de la longitud del intervalo deseada. Para ello, utilizaremos el método de la cantidad pivotal.

3.1.1. Una muestra

Sea x_i la respuesta del i -ésimo sujeto muestreado, $i = 1, 2, \dots, n$ (en investigación clínica, x_i podría ser la diferencia entre pares emparejados, como por ejemplo las respuestas entre antes y después del tratamiento). Supongamos que las x_i 's son variables aleatorias normales, independientes, y que están idénticamente distribuidas (i.i.d), con media μ y varianza σ^2 . Tenemos

$$x_i \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2) .$$

Tomamos

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{y} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 ,$$

la media y la varianza muestrales de las x_i 's, respectivamente.

En primer lugar, vamos a suponer el caso de que la varianza σ^2 es conocida. Tenemos que

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

es un pivote para la media μ . Sabemos que la función de densidad de Z es simétrica respecto del origen, y el intervalo más ajustado que buscamos lo obtenemos al tomar colas iguales. Por tanto, tomamos

$$P_\mu(Z < k_1) = \frac{\alpha}{2} ,$$

$$P_\mu(Z > k_2) = \frac{\alpha}{2} .$$

Tenemos entonces

$$k_1 = z_{1-\alpha/2} = -z_{\alpha/2} ,$$

$$k_2 = z_{\alpha/2} ,$$

donde " $z_{\alpha/2}$ " es el cuartil " $\alpha/2$ " superior de la distribución normal estándar. Así, mediante el método de la cantidad pivotal, obtenemos

$$P_{\mu} \left(-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \right) = 1 - \alpha .$$

Despejando μ , llegamos a

$$P_{\mu} \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha .$$

Por tanto, el intervalo de confianza que buscamos es

$$IC_{1-\alpha}(\mu) = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] .$$

Ahora, dado un n , sabemos que una mayor fiabilidad, es decir, un α menor, implica una menor precisión, por tanto, un intervalo de confianza más largo. Entonces, si queremos aumentar la probabilidad de acierto perderemos precisión en nuestra estimación. Por tanto, tenemos que tener controlados estos dos parámetros. Tomamos l la longitud del intervalo de confianza, y tenemos que:

$$l = \left(\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 2z_{\alpha/2} \frac{\sigma}{\sqrt{n}} .$$

Finalmente, despejando n , obtenemos la clásica fórmula utilizada por los autores deterministas:

$$n = \frac{4z_{\alpha/2}^2 \sigma^2}{l^2} . \quad (1)$$

Varianza σ^2 desconocida

Ahora, vamos a ver el caso en el que la varianza σ^2 es desconocida. En este caso, sustituimos σ^2 por la varianza muestral s , y tenemos que

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1} ,$$

donde t_{n-1} es una distribución *t de student* con $n - 1$ grados de libertad, es un pivote para μ .

Ahora, procediendo de una manera similar a la vista en el caso de varianza conocida, llegamos al intervalo de confianza que estábamos buscando:

$$IC_{1-\alpha}(\mu) = \left[\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right] .$$

Finalmente, utilizando la longitud del intervalo y el mismo razonamiento anterior, obtenemos la siguiente expresión:

$$n = \frac{4(t_{\alpha/2, n-1})^2 s^2}{l^2} . \quad (2)$$

Observación 3.1 Si nos fijamos, de esta expresión no podemos despejar explícitamente el valor de n y el valor de s^2 será desconocido a priori. Por ello, a la hora de obtener este tamaño muestral, los investigadores frecuentistas proponen realizar una estimación del parámetro σ^2 , basándose en información de estudios previos o realizando un experimento con una muestra pequeña, para posteriormente aplicar la fórmula (1).

3.1.2. Dos muestras

En investigación clínica muchas veces se realizan ensayos con el fin de evaluar la eficacia o la seguridad de un fármaco. Con este fin, se suelen comparar los resultados obtenidos entre grupos de personas con un control placebo o un agente de control activo como una terapia ya conocida; y grupos que están probando el nuevo tratamiento. Para ello, será interesante realizar una estimación de la diferencia de las medias de una variable determinada para los distintos grupos.

Sea x_{ij} la respuesta observada del sujeto j en el grupo de tratamiento i , con $j = 1, \dots, n_i$, $i = 1, 2$. Asumimos que cada x_{ij} son independientes y siguen una distribución normal, con media μ_i y varianza común σ^2 . Es decir,

$$x_{ij} \stackrel{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2) .$$

Ahora, sean

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad \text{y} \quad s^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 .$$

la media y la varianza muestral del grupo de tratamiento i , la idea será realizar una estimación de la diferencia de sus medias. Por tanto, vamos a tomar el parámetro $\theta = \mu_1 - \mu_2$.

Además, como estamos hablando de diferentes grupos de tratamiento, en la práctica es posible que nos encontremos en la situación de querer contar con distintos tamaños muestrales para cada uno de ellos. Esta diferencia deberá ser prefijada antes de realizar el estudio, es decir, se fijará un valor k para la diferencia proporcional entre estos tamaños muestrales. Esto es,

$$k = \frac{n_1}{n_2} .$$

Uno de los métodos más comunes en investigación clínica para hallar este valor de k se lleva a cabo teniendo en cuenta los costes que supondrán cada uno de los grupos. Para ello, se toma el siguiente ratio:

$$\frac{n_1}{n_2} = \sqrt{\frac{c_2}{c_1}} ,$$

donde c_1, c_2 representan los costes de cada grupo.

Varianza σ^2 conocida

Supongamos que la varianza σ es conocida y es común para ambas muestras. Realizando un procedimiento similar al visto anteriormente en la Sección 3.1.1, obtenemos el intervalo de confianza:

$$IC_{1-\alpha}(\theta) = \left[(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} , (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} \right] .$$

Ahora, tomando $n_1 = kn_2$, y despejando el valor de n_2 en función de la longitud del intervalo de confianza l de la misma manera que antes. Obtenemos el sistema:

$$\begin{cases} n_1 = kn_2 \\ n_2 = \frac{8(z_{\alpha/2})^2 \sigma^2}{l^2} \end{cases} . \quad (3)$$

Varianza σ^2 desconocida

Supongamos que la varianza σ no es conocida pero sí es común para ambas muestras. Tomaremos la varianza muestral s^2 y realizando un procedimiento similar al visto en el apartado para una muestra, obtenemos el intervalo de confianza:

$$IC_{1-\alpha}(\theta) = \left[(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2} \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} , (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2} \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} \right] .$$

Ahora, tomando $n_1 = kn_2$, y despejando el valor de n_2 en función de la longitud del intervalo de confianza l de la misma manera que antes. Obtenemos el sistema:

$$\begin{cases} n_1 &= kn_2 \\ n_2 &= \frac{8(t_{\alpha/2, n_2(1+k)-2})^2 s^2}{l^2} \end{cases} \quad (4)$$

Igual que pasaba con la fórmula para la estimación de una muestra con varianza desconocida (2), los tamaños muestrales no se pueden despejar explícitamente. Por ello, normalmente se trabaja igual que con una muestra; se obtiene una estimación de σ^2 y se utiliza la expresión (3).

3.1.3. Ejemplo

Ejemplo 3.2 *Vamos a tomar como ejemplo del experimento clínico propuesto en [5], que se basa en el artículo [14]. Como alternativa al uso de medicamentos, la idea del estudio será analizar un cambio en los hábitos alimenticios de los pacientes para combatir la hipertensión arterial. Concretamente, la idea es estudiar el efecto del consumo de leche agria en la presión sistólica; es decir, la presión que ejerce el corazón al bombear sangre. Para ello, buscaremos realizar una estimación de la media de esta presión arterial mediante un intervalo de confianza. Tomando un valor $\sigma^2 = 25$, en la Figura 1, podemos ver los resultados obtenidos a través de la fórmula frecuentista (1) para distintos valores de l y α .*

Tamaño Muestral para Estimación de una Media					
	alpha = 0.01	alpha = 0.03	alpha = 0.05	alpha = 0.07	alpha = 0.09
$l = 1$	166.00	118.00	97.00	83.00	72.00
$l = 2$	42.00	30.00	25.00	21.00	18.00
$l = 3$	19.00	14.00	11.00	10.00	8.00

Figura 1: Tamaños muestrales para la estimación de una media con $\sigma^2 = 25$.

3.2. Compación de medias

Hemos visto en la Sección 3.1 cómo obtener el tamaño muestral necesario para realizar una estimación de una diferencia de medias. A la hora de realizar un estudio, en lugar de cuál será el valor de esta diferencia de medias, podríamos preguntarnos directamente

si existe diferencia entre ellas, o si esta diferencia es mayor que cierto valor. Para ello, será interesante realizar contrastes de hipótesis, tanto para analizar si un tratamiento está funcionando o no, como para analizar si está teniendo el efecto que esperamos.

Para obtener el tamaño muestral necesario para realizar contrastes de hipótesis, los autores deterministas proponen lo que se conoce como un análisis de potencia. La idea principal será tomar un valor aceptable de α , e intentar trabajar con el valor de β . Vamos a resumir el razonamiento aplicado tomando como referencia la explicación propuesta en [12].

Partiremos de contrastes de hipótesis, donde el nivel de significancia α estará controlado bajo la hipótesis nula H_0 , mediante el test

$$P_{H_0}(T \in R_\alpha) = \alpha .$$

Analizaremos el caso concreto en el que nuestro estadístico T será continuo y obtendremos la región crítica del test mediante técnicas vistas en la asignatura de inferencia estadística, en este caso trabajaremos obteniendo un TUMP o aplicando el test de razón de verosimilitudes (TRV). Esta región crítica será de la forma $R_\alpha =]c_\alpha, \infty[$, con $c_\alpha \in \mathbb{R}$.

Por otro lado, la potencia estará controlada bajo la hipótesis alternativa H_1 y un valor de $1 - \beta$ será conseguido si

$$P_{H_1}(T \in R_\alpha) = 1 - \beta .$$

Ahora, dadas las funciones de distribución F_0 y F_1 , bajo H_0 y H_1 respectivamente, obtenemos el siguiente sistema de ecuaciones:

$$\left. \begin{aligned} F_0(c_\alpha) &= 1 - \alpha \\ F_1(c_\alpha) &= \beta \end{aligned} \right\} .$$

Entonces, despejando c_α de ambas expresiones e igualando obtenemos

$$F_1(F_0^{-1}(1 - \alpha)) = \beta .$$

Finalmente, la función F_1 (y normalmente también F_0) dependerá del tamaño muestral n , por lo que bastará despejarlo de esta expresión para obtener las fórmulas que buscamos.

En la Figura 2 podemos ver de manera visual las funciones de densidad bajo cada hipótesis y la relación que existe entre las funciones de distribución y las probabilidades de cometer los distintos errores.

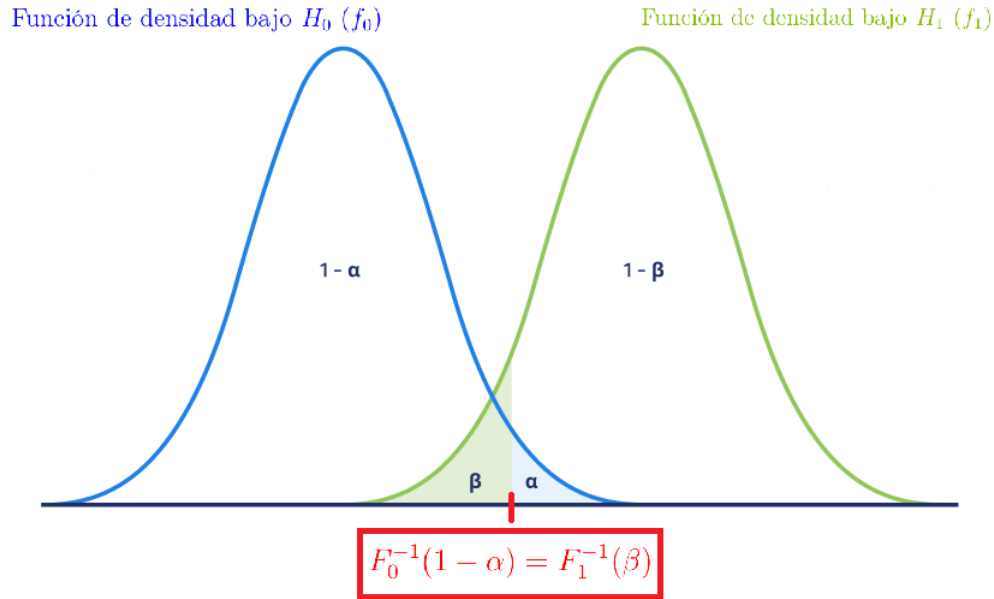


Figura 2: Funciones de densidad bajo la hipótesis nula (f_0) y bajo la hipótesis alternativa (f_1) junto con la probabilidad relacionada de cometer un error tipo I (α) y error tipo II (β).

3.2.1. Diseño con una muestra

En esta sección vamos a suponer que estamos trabajando con una única muestra de tamaño n , que será el valor que estamos buscando. En este tipo de ensayos clínicos se suele medir información de los individuos 2 veces a lo largo del estudio: una antes de recibir el tratamiento y la otra después, con el fin de evaluar si está surgiendo el efecto deseado.

Sea x_i la respuesta del i -ésimo sujeto muestreado, $i = 1, 2, \dots, n$. Supongamos que las x_i 's son variables aleatorias normales, independientes, y que están idénticamente distribuidas (i.i.d), con media μ y varianza σ^2 . Tenemos

$$x_i \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2) .$$

Además, definimos

$$\epsilon = \mu - \mu_0$$

como la diferencia entre la respuesta media de un fármaco de prueba (μ) y un valor de referencia (μ_0). Sin pérdida de generalidad, consideramos $\epsilon > 0$ como una indicación de mejora del fármaco de prueba en comparación con el valor de referencia. De la misma manera, consideramos $\epsilon < 0$ una indicación de empeoramiento del fármaco respecto del valor de referencia.

3.2.1.1. Test de Igualdad

Para evaluar si existen diferencias entre las medias, consideramos el siguiente contraste de hipótesis:

$$\left. \begin{array}{l} H_0 : \epsilon = 0 \\ H_1 : \epsilon \neq 0 \end{array} \right\} .$$

Varianza σ^2 conocida

Supongamos que σ^2 es conocida. Vamos a aplicar el test de razón de verosimilitudes. Como las x'_i s siguen una distribución normal, la función de verosimilitud será

$$L_{(x_1, \dots, x_n)}(\mu) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} .$$

Ahora, como $\Theta = \mathbb{R}$, $\Theta_0 = \mu_0$ y \bar{x} es un EMV para μ , entonces

$$\sup_{\mu \in \Theta_0} L_{(x_1, \dots, x_n)}(\mu) = L_{(x_1, \dots, x_n)}(\mu_0) ,$$

$$\sup_{\mu \in \Theta} L_{(x_1, \dots, x_n)}(\mu) = L_{(x_1, \dots, x_n)}(\hat{\mu}) = L_{(x_1, \dots, x_n)}(\bar{x}) .$$

Dividiendo, obtenemos la razón de verosimilitudes, que cumplirá:

$$\lambda(x_1, \dots, x_n) = \frac{L_{(x_1, \dots, x_n)}(\mu_0)}{L_{(x_1, \dots, x_n)}(\bar{x})} = \frac{\exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2 \right\}}{\exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}} < k .$$

Ahora, despejando los parámetros de interés, llegamos a

$$|\bar{x} - \mu_0| > K^* .$$

Por tanto, obtenemos el siguiente test:

$$\chi_1^* = \{(x_1, \dots, x_n) : |\bar{x} - \mu_0| > K^*\} .$$

Hallamos el valor de K^* tomando:

$$P_{\mu_0} \left(|z| > \frac{K^*}{\sigma/\sqrt{n}} \right) = \alpha .$$

Y, por la simetría de la distribución normal, esto es equivalente a

$$P_{\mu_0} \left(z > \frac{K^*}{\sigma/\sqrt{n}} \right) = \frac{\alpha}{2} .$$

Finalmente, despejando K^* y sustituyendo en el test, obtenemos:

$$\chi_1^* = \left\{ (x_1, \dots, x_n) : \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2} \right\}.$$

Esto quiere decir que rechazaremos la hipótesis nula con nivel de significancia α si

$$\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2},$$

Ahora, bajo la hipótesis alternativa $H_1 : \epsilon \neq 0$, la potencia del test está determinada por

$$\Phi\left(\frac{\sqrt{n}\epsilon}{\sigma} - z_{\alpha/2}\right) + \Phi\left(-\frac{\sqrt{n}\epsilon}{\sigma} - z_{\alpha/2}\right) = 1 - \beta,$$

donde Φ es la función de distribución acumulada de una normal estándar. Ignorando un pequeño valor $\leq \alpha/2$, la potencia es aproximadamente

$$\Phi\left(\frac{\sqrt{n}|\epsilon|}{\sigma} - z_{\alpha/2}\right) = 1 - \beta.$$

Como resultado, obtendremos el tamaño muestral necesario para conseguir una potencia de $1 - \beta$ resolviendo la ecuación

$$\frac{\sqrt{n}|\epsilon|}{\sigma} - z_{\alpha/2} = z_{\beta}.$$

Despejando, llegamos a

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\epsilon^2}. \quad (5)$$

Varianza σ^2 desconocida

Si σ^2 no es conocido, entonces puede ser reemplazado por la varianza muestral s^2 . En este caso, realizando el test de razón de verosimilitudes mediante un proceso similar al anteriormente visto para el caso en el que la varianza es conocida, rechazamos la hipótesis nula H_0 si

$$\left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| > t_{\alpha/2, n-1},$$

Bajo la hipótesis alternativa ($\epsilon \neq 0$), la potencia del test viene dada por

$$1 - T_{n-1}\left(t_{\alpha/2, n-1} \left| \frac{\sqrt{n}\epsilon}{\sigma} \right| \right) + T_{n-1}\left(-t_{\alpha/2, n-1} \left| \frac{\sqrt{n}\epsilon}{\sigma} \right| \right) = 1 - \beta,$$

donde $T_{n-1}(\cdot|\theta)$ es la función de distribución acumulada de una *t de student* no centrada con $n - 1$ grados de libertad y con parámetro de no centralización θ .

Ignorando un valor pequeño $\leq \alpha/2$, la potencia $1 - \beta$ es aproximadamente

$$1 - T_{n-1} \left(t_{\alpha/2, n-1} \left| \frac{\sqrt{n}|\epsilon|}{\sigma} \right| \right) = 1 - \beta .$$

Por tanto, podemos obtener el tamaño muestral resolviendo la ecuación

$$T_{n-1} \left(t_{\alpha/2, n-1} \left| \frac{\sqrt{n}|\epsilon|}{\sigma} \right| \right) = \beta. \quad (6)$$

Observación 3.3 Cuando n es suficientemente grande, $t_{\alpha/2, n-1} \approx z_{\alpha/2}$ y $t_{\beta, n-1} \approx z_{\beta/2}$. Por lo tanto, se cumple que

$$T_{n-1} \left(t_{\alpha/2, n-1} \left| \frac{\sqrt{n}|\epsilon|}{\sigma} \right| \right) \approx \Phi \left(z_{\alpha/2} - \frac{\sqrt{n}\epsilon}{\sigma} \right) .$$

Podríamos entonces utilizar la fórmula (5) también en el caso de σ desconocida.

Observación 3.4 Normalmente, cuando la varianza σ^2 es desconocida, del mismo modo que hemos visto en la Observación 3.1, los investigadores trabajan obteniendo una estimación de este parámetro a través de información de estudios previos o realizando un ensayo con un grupo muy reducido de personas, y directamente utilizan las fórmulas para la varianza conocida.

3.2.1.2. Test de no inferioridad o superioridad

El problema de no inferioridad o superioridad se puede resolver mediante el contraste de hipótesis:

$$\left. \begin{array}{l} H_0 : \epsilon \leq \delta \\ H_1 : \epsilon > \delta \end{array} \right\} ,$$

donde δ es el margen de superioridad o no inferioridad. Cuando $\delta > 0$, el rechazo de la hipótesis nula H_0 indica superioridad de la nueva media (μ) respecto del valor de referencia (μ_0). En caso contrario, cuando $\delta < 0$, el rechazo de la hipótesis nula implica no inferioridad.

Varianza σ^2 conocida

Supongamos que σ^2 es conocido, en este caso vamos a buscar un TUMP aplicando el Teorema (2.2). Para ello, necesitamos encontrar un estimador suficiente para μ respecto

del cual se cumpla la razón de verosimilitudes. Por tanto, debido a que las x'_i s siguen una distribución normal, vamos a tomar la función de densidad de una distribución normal. Observamos que podemos expresarla de la forma:

$$f_{\mu}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\} = a(\mu)b(x_1, \dots, x_n) \exp \left\{ \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i \right\},$$

con $a(\mu) > 0$ y $b(x_1, \dots, x_n) > 0$. Por tanto, tenemos que x pertenece a la familia exponencial uniparamétrica. Además, tomando

$$T = \bar{x} \quad \text{y} \quad q(\mu) = \frac{n\mu}{\sigma^2},$$

tenemos que $q(\mu)$ estrictamente creciente. Entonces, por la Proposición (2.1), podemos afirmar que se cumple la RVM respecto de T .

Por tanto, aplicando ahora el Teorema de Karlin-Rubin (2.2), un TUMP será

$$\chi_1^* = \{(x_1, \dots, x_n) / \bar{x} > K\}.$$

Finalmente, obtenemos el valor de K siguiendo un razonamiento similar al visto en la Sección 3.2.1.1, y, despejando, llegamos al test:

$$\chi_1^* = \left\{ (x_1, \dots, x_n) : \frac{\bar{x} - \mu_0 - \delta}{\sigma/\sqrt{n}} > z_{\alpha} \right\}.$$

Esto quiere decir que rechazamos H_0 con nivel de significancia α si

$$\frac{\bar{x} - \mu_0 - \delta}{\sigma/\sqrt{n}} > z_{\alpha}.$$

Bajo la hipótesis alternativa $\epsilon > \delta$, tenemos que la potencia del test es

$$\Phi \left(\frac{\sqrt{n}(\epsilon - \delta)}{\sigma} - z_{\alpha} \right) = 1 - \beta.$$

Como resultado, podremos obtener el tamaño muestral necesario para una potencia de $1 - \beta$ resolviendo

$$\frac{\sqrt{n}(\epsilon - \delta)}{\sigma} - z_{\alpha} = z_{\beta}.$$

Despejando n , obtenemos

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{(\epsilon - \delta)^2}. \quad (7)$$

Varianza σ^2 desconocida

Cuando σ^2 no es conocido, podemos reemplazarlo por s^2 . Procediendo de una manera similar al apartado anterior, obtenemos entonces que H_0 será rechazado con un nivel de significancia α si

$$\frac{\bar{x} - \mu_0 - \delta}{s/\sqrt{n}} > t_{\alpha, n-1} .$$

Por tanto, la potencia del test viene dada por

$$1 - T_{n-1} \left(t_{\alpha, n-1} \left| \frac{\sqrt{n}(\epsilon - \delta)}{\sigma} \right| \right) = 1 - \beta .$$

Finalmente, obtenemos el tamaño muestral necesario para conseguir una potencia $1 - \beta$ resolviendo

$$T_{n-1} \left(t_{\alpha, n-1} \left| \frac{\sqrt{n}(\epsilon - \delta)}{\sigma} \right| \right) = \beta . \quad (8)$$

Aplicando la aproximación vista en la Observación 3.3, cuando n es suficientemente grande, podemos aproximar la fórmula (8) a partir de (7). Así es como se suele proceder.

3.2.2. Diseño paralelo con dos muestras

Sea x_{ij} la respuesta observada del sujeto j en el grupo de tratamiento i , con $j = 1, \dots, n_i$, $i = 1, 2$. Asumimos que cada x_{ij} son independientes y siguen una distribución normal, con media μ_i y varianza σ^2 . Es decir,

$$x_{ij} \stackrel{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2).$$

Ahora, sean

$$\bar{x}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad \text{y} \quad s^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\cdot})^2$$

la media y la varianza muestral del grupo de tratamiento i .

Además, vamos a tomar $\epsilon = \mu_2 - \mu_1$ la diferencia de medias de la variable a estudiar entre el grupo de test (μ_2) y el grupo placebo o de control (μ_1). Sin pérdida de generalidad, vamos a considerar $\epsilon > 0$ como un indicador de mejora del nuevo fármaco o producto a probar con respecto al placebo o medicamento que ya teníamos.

Al tratar con dos grupos de tratamiento, puede darse la situación de querer contar con tamaños muestrales distintos. Por tanto, de la misma manera que en la Sección 3.1.2, vamos a tomar

$$k = \frac{n_1}{n_2} .$$

3.2.2.1. Test de Igualdad

El objetivo de este test será estudiar si existe diferencia de medias entre ambos grupos. Para ello, consideramos el contraste de hipótesis

$$\left. \begin{array}{l} H_0 : \epsilon = 0 \\ H_1 : \epsilon \neq 0 \end{array} \right\} .$$

Varianza σ^2 conocida

Si σ^2 es conocida, aplicando un test de razón de verosimilitudes, llegamos a que rechazaremos H_0 con un nivel de significancia α si

$$\left| \frac{\bar{x}_{1.} - \bar{x}_{2.}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| > z_{\alpha/2} .$$

Por tanto, bajo la hipótesis alternativa $\epsilon \neq 0$, la potencia sobre el test es

$$\Phi \left(\frac{\epsilon}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - z_{\alpha/2} \right) + \Phi \left(-\frac{\epsilon}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - z_{\alpha/2} \right) = 1 - \beta .$$

Ignorando un valor pequeño de $\neq \alpha/2$, podemos aproximar esta expresión a

$$\Phi \left(\frac{|\epsilon|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - z_{\alpha/2} \right) = 1 - \beta .$$

Finalmente, el tamaño muestral necesario para conseguir una potencia de $1 - \beta$ puede ser obtenido resolviendo la ecuación

$$\frac{|\epsilon|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - z_{\alpha/2} = z_{\beta} . \quad (9)$$

Despejando, obtenemos los tamaños muestrales resolviendo el sistema:

$$\begin{cases} n_1 = kn_2 \\ n_2 = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2 (1 + \frac{1}{k})}{\epsilon^2} \end{cases} . \quad (10)$$

Varianza σ^2 desconocida

Por otro lado, supongamos ahora que σ^2 es desconocida. La idea será la misma que cuando trabajábamos con una sola muestra, reemplazamos σ^2 por la varianza muestral s^2 . Ahora, aplicando un test de razón de verosimilitudes, llegamos a que la hipótesis nula H_0 será rechazada si

$$\left| \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| > t_{\alpha/2, n_1+n_2-2} ,$$

Por tanto, bajo la hipótesis alternativa $\epsilon \neq 0$, la potencia del test viene dado por

$$1 - T_{n_1+n_2-2} \left(t_{\alpha/2, n_1+n_2-2} \left| \frac{\epsilon}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \right) + T_{n_1+n_2-2} \left(-t_{\alpha/2, n_1+n_2-2} \left| \frac{\epsilon}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \right) .$$

Ignorando un valor pequeño $\leq \alpha/2$, la potencia es aproximadamente

$$1 - T_{n_1+n_2-2} \left(t_{\alpha/2, n_1+n_2-2} \left| \frac{|\epsilon|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \right) = 1 - \beta .$$

Finalmente, tomando $n_1 = kn_2$ y sustituyendo en la fórmula, obtenemos el valor de n_2 , resolviendo la ecuación

$$T_{(1+k)n_2-2} \left(t_{\alpha/2, (1+k)n_2-2} \left| \frac{\sqrt{n_2}|\epsilon|}{\sigma \sqrt{1 + \frac{1}{k}}} \right| \right) = \beta . \quad (11)$$

Del mismo modo que en el apartado anterior, aplicando la aproximación vista en la Observación 3.3, podemos simplificar la expresión (11) aproximándola a la fórmula para la varianza conocida (10) cuando n_1, n_2 son suficientemente grandes.

3.2.2.2. Test de no inferioridad o superioridad

Analizaremos la superioridad o no inferioridad de la nueva media resolviendo el contraste de hipótesis:

$$\left. \begin{array}{l} H_0 : \epsilon \leq \delta \\ H_1 : \epsilon > \delta \end{array} \right\} .$$

donde δ es el margen de superioridad o no inferioridad. Cuando $\delta > 0$, el rechazo de la hipótesis nula H_0 indica la superioridad del grupo del medicamento que está siendo testado por encima del grupo placebo o de control. Cuando $\delta < 0$, el rechazo de esta hipótesis nula indica la no inferioridad del test respecto del control.

Varianza σ^2 conocida

Cuando σ^2 es conocida, obteniendo un TUMP y aplicando un razonamiento similar al caso de una muestra, llegamos a que la hipótesis nula H_0 será rechazada con nivel de significancia α si

$$\frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > z_\alpha .$$

Bajo la hipótesis alternativa de $\epsilon > \delta$, la potencia bajo el test viene dada por

$$\Phi \left(\frac{\epsilon - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - z_\alpha \right) = 1 - \beta .$$

Por tanto, los tamaños muestrales necesarios para una potencia de $1 - \beta$ pueden ser obtenidos resolviendo

$$\frac{\epsilon - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - z_\alpha = z_\beta .$$

Más concretamente, obtenemos n_1 y n_2 resolviendo:

$$\begin{cases} n_1 = kn_2 \\ n_2 = \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2 (1 + \frac{1}{k})}{(\epsilon - \delta)^2} \end{cases} . \quad (12)$$

Varianza σ^2 desconocida

Por otro lado, cuando σ^2 no es conocido, podemos reemplazarlo por s^2 . Ahora, procediendo de la misma manera que en los casos anteriores, la hipótesis nula H_0 será rechazada con nivel de significancia α si

$$\frac{\bar{x}_1 - \bar{x}_2 - \delta}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{\alpha, n_1+n_2-2} .$$

Bajo la hipótesis alternativa de $\epsilon > \delta$, la potencia obtenida en el test viene dado por

$$1 - T_{n_1+n_2-2} \left(t_{\alpha, n_1+n_2-2} \left| \frac{\epsilon - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \right) = 1 - \beta .$$

Por tanto, el tamaño muestral necesario para alcanzar una potencia de $1 - \beta$ será obtenido resolviendo

$$T_{n_1+n_2-2} \left(t_{\alpha, n_1+n_2-2} \left| \frac{\epsilon - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \right) = \beta . \quad (13)$$

Aplicando la aproximación vista en la Observación 3.3, cuando n_1, n_2 son suficientemente grandes, podemos resolver el sistema de ecuaciones para varianza conocida (12) con el fin de obtener una aproximación de (13).

3.2.3. Ejemplo

Ejemplo 3.5 *Tomamos los datos vistos en el Ejemplo 3.1.3. Supongamos ahora que queremos realizar un estudio con el mismo fin, analizar el efecto del consumo de leche agria sobre la reducción de la hipertensión arterial, pero esta vez queremos analizar las posibles diferencias entre dos grupos de pacientes: un grupo con gente que consume leche agria a lo largo del estudio; y, otro grupo placebo. Para ello, queremos analizar si habrá una diferencia entre las medias de ambos grupos a través de un contraste de hipótesis de no inferioridad o superioridad. Concretamente, vamos a analizar si las personas que toman leche agria tienen de media $\delta = 2\text{mmHg}$ menos de presión sistólica. Vamos entonces a suponer que ambos grupos siguen una distribución normal con $\sigma^2 = 25$ común, y buscaremos dos grupos con el mismo número de personas, es decir, $n_1 = n_2$. En la Figura 3, podemos observar los tamaños muestrales obtenidos para distintos valores de α y β .*

Tamaño Muestral para diferentes valores de alpha y beta

	beta = 0.24	beta = 0.22	beta = 0.2	beta = 0.18	beta = 0.16	beta = 0.14	beta = 0.12	beta = 0.1
alpha = 0.01	136.32	141.78	147.65	154.02	161.00	168.76	177.52	187.66
alpha = 0.02	116.32	121.37	126.81	132.72	139.21	146.43	154.60	164.07
alpha = 0.03	104.61	109.40	114.57	120.19	126.37	133.25	141.05	150.11
alpha = 0.04	96.29	100.89	105.85	111.26	117.21	123.84	131.37	140.11
alpha = 0.05	89.83	94.28	99.08	104.31	110.08	116.51	123.82	132.31
alpha = 0.06	84.56	88.87	93.54	98.62	104.23	110.49	117.61	125.90
alpha = 0.07	80.10	84.30	88.84	93.80	99.27	105.39	112.35	120.45
alpha = 0.08	76.23	80.34	84.77	89.62	94.97	100.96	107.77	115.70
alpha = 0.09	72.83	76.84	81.18	85.93	91.17	97.03	103.71	111.50
alpha = 0.1	69.78	73.71	77.96	82.62	87.76	93.51	100.08	107.73

Figura 3: Tamaños muestrales para la comparación de dos medias con $\epsilon = 0$, $\delta = 2$.

El caso más utilizado por los investigadores es tomar los valores $\alpha = 0,05$ y $\beta = 0,2$. Para realizar un estudio con estas condiciones necesitaremos contar con al menos 100 individuos en cada grupo.

Observación 3.6 *En [13] se pueden encontrar las tablas completas para todos los métodos frecuentistas estudiados, con los n resultantes para todos los valores de α y β posibles.*

4. Métodos bayesianos

Recordamos la notación y resultados vistos en la Sección 2.2. Sea $\theta \in \Theta$, un parámetro genérico. Supongamos que la información previa del valor θ es descrita a través de una distribución previa caracterizada por $f(\theta)$, que será la distribución a priori. Consideremos $x = (x_1, \dots, x_n)$, donde x_i son independientes y son idénticamente distribuidas aleatoriamente. Supongamos que $x \in S$, con S el espacio muestral asociado. Entonces, la distribución marginal de x viene dada por

$$f(x) = \int_{\Theta} f(x|\theta)f(\theta)d\theta ,$$

donde $f(x|\theta)$ es la distribución condicional de x dada por el parámetro θ . Entonces, dado x , aplicando el Teorema 2.4, la distribución posterior de θ viene dada por

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} .$$

4.1. Motivación

En la Sección 3.1 hemos visto algunas fórmulas clásicas basadas en la perspectiva frecuentista de la probabilidad. En [9] se tratan varios problemas que pueden aparecer a la hora de trabajar con estos métodos.

Por un lado, en la práctica, la estadística frecuentista se realiza basándose exclusivamente en los datos observados al final del estudio. Por ello, para los investigadores, desde esta perspectiva, es difícil planificar el estudio en un inicio. De hecho, como hemos visto en la Observación 3.1, cuando el valor de σ^2 sea desconocido, se verán obligados a realizar predicciones basadas en información de estudios previos o en opiniones subjetivas para estimar este parámetro. Además, este valor predicho juega un papel fundamental en la mayoría de fórmulas, por lo que el tamaño muestral resultante puede ser excesivamente grande o pequeño si cometemos un error. Por tanto, es interesante replantear el problema desde un punto de vista bayesiano, ya que la subjetividad se tendrá en cuenta desde un inicio y será utilizada de una manera más clara y explícita.

Por otro lado, existen situaciones en las que podemos conocer información inicial relacionada con las medias a estudiar, que puede ser incluida como información a priori si trabajamos desde el punto de vista bayesiano. Además, pueden darse situaciones trabajando a partir de las fórmulas clásicas, en las que el hecho de ignorar esta información, nos hará tomar un tamaño muestral innecesariamente grande, el cual puede provocar pérdidas económicas u ocasionar problemas de salud, en el caso de estudios clínicos.

Estas limitaciones de las fórmulas frecuentistas motivaron a los autores Joseph y Belisle [9] a analizar el problema del tamaño muestral desde una visión basada en la estadística bayesiana. Concretamente, fueron capaces de proporcionar tres criterios diferentes para estudios cuyo fin es la una estimación de un parámetro mediante intervalos de credibilidad posterior. Vamos a estudiar cómo funcionan y a realizar un comparación entre ellos.

Observación 4.1 *Cuando usamos fórmula frecuentistas como (1), es suficiente con especificar los valores de l , α y el valor de σ . Sin embargo, a la hora de realizar la estimación bayesiana, conocer l , α y la distribución previa de θ no será suficiente, ya que se requiere conocer la distribución que seguirán los datos de la muestra, su verosimilitud, cosa que evidentemente es desconocida al inicio del estudio. Consideraremos una predicción de esta verosimilitud previa para poder aplicar el Teorema de Bayes.*

4.2. 3 criterios de selección

4.2.1. Average Coverage Criterion (ACC)

Fijando la longitud del intervalo de credibilidad posterior l , la probabilidad de cobertura $1 - \alpha$ variará en función de x a lo largo de S . El objetivo que se persigue con el criterio ACC es que los intervalos de credibilidad posterior tengan una alta probabilidad de contener el verdadero valor del parámetro. Para ello, trataremos de encontrar la n más pequeña que satisfaga que la media de cobertura de un intervalo de credibilidad posterior de longitud l , ponderado con $f(x)$, sea al menos de $1 - \alpha$. Es decir, que la media de las probabilidades de que el parámetro se encuentre dentro nuestros intervalos de credibilidad posterior de longitud l sea al menos $1 - \alpha$. Analíticamente, la menor n será obtenida resolviendo la inecuación

$$\int_S \left\{ \int_a^{a+l} f(\theta|x, n) d\theta \right\} f(x) dx \geq 1 - \alpha, \quad (14)$$

donde $a(x, n)$ es un estadístico que será determinado por los datos, y depende de x y de la longitud de la muestra. Diversos autores han propuesto soluciones para tratar con este estadístico. En primer lugar, en [1] se propone escoger un intervalo $(a, a + l)$ que sea simétrico respecto de la media. Por otro lado, en [10] se selecciona $(a, a + l)$ el mayor intervalo de densidad posterior (HDPI). En el caso de nuestro estudio, como trataremos únicamente con distribuciones normales, al ser simétricas, llegaremos a la misma solución aplicando cualquiera de estas dos ideas.

4.2.2. Average Length Criterion (ALC)

En el anterior criterio fijábamos la longitud del intervalo de credibilidad posterior l y optimizábamos el tamaño muestral intentando conseguir una probabilidad de al menos $1 - \alpha$ de encontrar el parámetro. Siguiendo una idea similar, esta vez fijaremos el nivel de probabilidad y seleccionaremos el tamaño muestral para el cual el intervalo tiene una longitud promedio l , que variará en función de los valores de $x \in S$. La idea entonces será buscar la l más pequeña posible; es decir, minimizar la longitud de estos intervalos de credibilidad, ya que cuánto más cortos sean, más precisas serán las estimaciones.

Analíticamente, ALC selecciona el menor tamaño muestral tal que

$$\int_S l'(x, n) f(x) dx \leq l, \quad (15)$$

donde $l'(x, n)$ es la longitud del intervalo de credibilidad posterior al $100(1 - \alpha)\%$ para cada valor $x \in S$, que será determinado resolviendo

$$\int_a^{a+l'(x, n)} f(\theta|x, n) d\theta = 1 - \alpha. \quad (16)$$

De la misma manera que en el criterio anterior, se aplican los mismos métodos para obtener el valor del estadístico a .

4.2.3. Worst Outcome Criterion (WOC)

En la práctica, los investigadores pueden no estar satisfechos con los criterios anteriores, debido a que éstos se basan en garantizar que el promedio de las probabilidades o de las longitudes, fijando el otro parámetro, cumplirán ciertas cosas. Por ejemplo, el tamaño muestral seleccionado con ALC solo asegura que el promedio de la longitud del intervalo de credibilidad posterior no será superior a l , pero podemos tener una gran cantidad de intervalos que superen este valor. Entonces, se pensó que un investigador, a la hora de hacer un estudio, desearía seleccionar el tamaño muestral tal que la predicción de la longitud del intervalo de credibilidad posterior sea lo más cercana a l como sea posible, y que no sea mayor que esta en ningún caso.

En esta situación, el siguiente criterio (WOC) resultará de utilidad. Este método pretende maximizar la probabilidad de que un intervalo de credibilidad cubra el verdadero valor del intervalo, dentro de un margen de error establecido. Analíticamente, selecciona el menor tamaño muestral resolviendo la inecuación

$$\inf_{x \in S_0} \left\{ \int_a^{a+l} f(\theta|x, n) d\theta \right\} f(x) dx \geq 1 - \alpha, \quad (17)$$

donde S_0 es un subespacio seleccionado apropiadamente del espacio original S . Por ejemplo, podríamos considerar S_0 como la región que contiene el 95 % de los $x \in S$ más probables. En este caso, WOC aseguraría que hay una garantía del 95 % de que la longitud del intervalo de credibilidad posterior de credibilidad $100(1 - \alpha) \%$ sea menor que l . Del mismo modo que en los criterios anteriores, se utilizan los métodos mencionados para obtener este valor a .

4.3. Estimación de una media. Criterios y comparación

A lo largo de esta sección consideramos vector de observaciones $x = (x_1, \dots, x_n)$, donde los x_i están idénticamente distribuidos y siguen una distribución normal $N(\mu, \sigma^2)$, con μ desconocidas. Además, para hacer uso de una notación más sencilla, vamos a dejar nuestras expresiones en función de la precisión τ .

Además, vamos a suponer que, a priori, las variables desconocidas μ y τ siguen las distribuciones

$$\begin{aligned}\tau &\sim \Gamma(v, \rho) , \\ \mu|\tau &\sim N(\mu_0, n_0\tau) .\end{aligned}$$

Las distribuciones de probabilidad que vamos a utilizar a lo largo de esta sección pueden encontrarse en el primer apéndice de [3]. Por otro lado, el detalle de los desarrollos de todos criterios que vamos a analizar se puede encontrar en los apéndices del artículo [9].

4.3.1. Precisión conocida

Supongamos que la precisión τ es conocida *a priori*. Vamos a buscar ahora cual será a posteriori la distribución que seguirá la media μ .

Siguiendo la notación de los autores bayesianos. En primer lugar, vamos a tomar la verosimilitud:

$$f(x_1, \dots, x_n|\mu) = \prod_{i=1}^n \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp \left\{ \frac{-\tau(x_i - \mu)^2}{2} \right\} = \left(\frac{\tau}{2\pi} \right)^{n/2} \exp \left\{ \frac{-\tau}{2} \sum_{i=1}^n (x_i - \mu)^2 \right\} .$$

Además, tomamos la función de densidad de la media a priori:

$$f(\mu) = \sqrt{\frac{n_0\tau}{2\pi}} \exp \left\{ \frac{-1}{2} n_0\tau(\mu - \mu_0)^2 \right\} .$$

Ahora, vamos a aplicar el Teorema de Bayes para obtener la distribución a posteriori:

$$f(\mu|x, \tau) = \frac{f(\mu)f(x|\mu, \tau)}{\int_{-\infty}^{\infty} f(\mu)f(x|\mu, \tau)d\mu} \propto \exp \left\{ -\frac{n_0\tau(\mu - \mu_0)^2}{2} - \frac{\tau \sum_{i=1}^n (x_i - \mu)^2}{2} \right\}.$$

Finalmente, si seguimos desarrollando equivalencias suponiendo que las distribuciones son conjugadas, con la idea entonces de que la resultante debe seguir una distribución normal. Llegamos a que nuestra función a posteriori cumple:

$$f(\mu|x, \tau) \propto \exp \left\{ \frac{-1}{2}(n + n_0)\tau \left(\mu - \frac{n_0\mu_0 + n\bar{x}}{n_0 + n} \right)^2 \right\}.$$

Esto se corresponde con la parte no constante de una distribución normal de la forma:

$$\mu|x \sim N(\mu_n, \tau_n),$$

donde

$$\begin{aligned} \tau_n &= (n + n_0)\tau, \\ \mu_n &= \frac{n_0\mu_0 + n\bar{x}}{n_0 + n}. \end{aligned}$$

Si nos fijamos, la precisión posterior τ_n solamente depende del tamaño muestral n y es independiente de los valores observados x . Por tanto, desarrollando los 3 criterios obtendremos la misma solución:

$$n \geq \frac{4z_{1-\alpha/2}^2}{\tau l^2} - n_0. \quad (18)$$

Esta solución es igual que la obtenida en [1], donde se aplican también otras técnicas basadas en la estadística bayesiana.

Observación 4.2 *Si el tamaño muestral a priori es $n_0 = 0$, entonces esta fórmula (18) se reduce a la del modelo frecuentista (1). Además, si las comparamos entre sí, podemos observar que el resultado obtenido con la aproximación bayesiana nunca va a ser mayor que el obtenido con la frecuentista.*

4.3.2. Precisión desconocida

Supongamos ahora que la precisión τ no es conocida. Procediendo ahora de una manera similar al apartado anterior, esta vez tomando (μ, τ) como variable conjunta de interés; y,

posteriormente, obteniendo la distribución marginal posterior de μ integrando la densidad conjunta obtenida con respecto a τ como hemos visto en la Sección 2.2, llegaremos a la distribución marginal de μ . Basándonos en los cálculos realizados en [6], obtenemos que esta distribución marginal viene dada por

$$\mu|x \sim t_{2v+n} \sqrt{\frac{\rho_n}{(n+n_0)\left(v+\frac{n}{s}\right)}} + \mu_n ,$$

donde

$$\begin{aligned} \mu &= \frac{n_0\mu_0 + n\bar{x}}{n+n_0} , \\ \rho_n &= \rho + \frac{1}{2}ns^2 + \frac{nn_0}{2(n+n_0)}(\bar{x} - \mu_0)^2 . \end{aligned}$$

En esta situación, al contrario que en el caso anterior, podemos ver que la precisión posterior varía en función de los valores de x . Por tanto, con cada criterio de selección obtendremos tamaños muestrales diferentes.

Average Coverage Criterion (ACC)

En el artículo [1] se demuestra que el tamaño muestral con el método ACC viene dado por

$$n = \frac{4\rho}{vl^2} t_{2v,1-\alpha/2}^2 - n_0 . \quad (19)$$

Observamos que v/ρ es la media a priori para la precisión τ , por tanto, esta fórmula del tamaño muestral para la precisión desconocida es bastante similar a la de precisión conocida (18); simplemente habría que sustituir los valores de τ por su media y el cuartil de una distribución normal $z_{1-\alpha/2}$ por el cuartil de una $t_{2v,1-\alpha}$. Si embargo, debido a que los grados de libertad de la distribución *t de student* no aumentan con el tamaño muestral, la ecuación (19) puede llevarnos a tamaños muestrales que sean muy diferentes de los obtenidos con las expresiones (1) y (18).

Average Length Criterion (ALC)

Basándonos en los cálculos realizados en [9], cuando n es suficientemente grande, el tamaño muestral requerido por el criterio ALC puede ser obtenido resolviendo la inecuación

$$2t_{n+2v,1-\alpha/2} \sqrt{\frac{2\rho}{(n+2v)(n+n_0)}} \frac{\Gamma\left(\frac{n+2v}{2}\right) \Gamma\left(\frac{2v-1}{2}\right)}{\Gamma\left(\frac{n+2v-1}{2}\right) \Gamma(v)} \leq l . \quad (20)$$

Desafortunadamente, no existe una solución explícita de esta inecuación. Sin embargo, hay algoritmos diseñados para estimar este tamaño muestral computacionalmente.

Por otro lado, en el caso de que n sea suficientemente grande ($n \rightarrow \infty$), podemos encontrar aproximaciones de esta fórmula; concretamente, vamos a ver la aproximación propuesta en [5] .

Aplicando la aproximación de Stirling a un cociente de gammas, tenemos

$$\frac{\Gamma(z+a)}{\Gamma(z+b)} = z^{a-b} \left(1 + \frac{(a-b)(a+b-1)}{z} + o(z^{-2}) \right) .$$

Sustituyendo por esta expresión en la división de gammas de la fórmula (20) que contienen n , obtenemos

$$\frac{\Gamma\left(\frac{n+2v}{2}\right)}{\Gamma\left(\frac{n+2v-1}{2}\right)} = \sqrt{\frac{n+2v}{2}} + o((n+2v)^{-1/2}) .$$

Finalmente, sustituyendo en nuestra fórmula (20), obtenemos la aproximación

$$n \geq \frac{4t_{n+2v,1-\alpha/2}^2 \rho}{l^2 v C_v} - n_0 , \quad (21)$$

donde

$$C_v = \left(\frac{\Gamma(v)}{\Gamma\left(\frac{2v-1}{2}\right)} \right)^2 \frac{1}{v} . \quad (22)$$

Worst Outcome Criterion (WOC)

Sea $S_0 \subset S$ el subespacio del espacio muestral tal que

$$\int_{S_0} f(x) dx = 1 - w ,$$

para una probabilidad $w > 0$. supongamos que $f(x) \leq f(y)$ para cualquier $x \in S, y \notin S$. De este modo, tomando los cálculos realizados en [9], el tamaño muestral necesario puede ser aproximado mediante la fórmula

$$\frac{l^2(n+2v)(n+n_0)}{8\rho \left(1 + \left(\frac{n}{2v}\right) F_{n,2v,1-w}\right)} \leq t_{n+2v,1-\alpha/2}^2 , \quad (23)$$

donde $F_{d_1,d_2,1-\alpha}$ es el cuartil $1 - \alpha$ de la distribución F de *Snedecor* con (d_1, d_2) grados de libertad.

Observación 4.3 *El subespacio S_0 no puede ser exactamente S , ya que tendríamos $w = 0$, y, por tanto, $F_{d_1, d_2, 1} = \infty$. En este caso el tamaño muestral no estaría bien definido.*

Igual que con el criterio anterior, cuando n es suficientemente grande ($n \rightarrow \infty$), podemos encontrar aproximaciones con tratamientos más sencillos para esta fórmula. De nuevo, vamos a tomar como ejemplo la propuesta en [5].

En primer lugar, tomamos las aproximaciones

$$z_{1-\alpha/2} \approx t_{n+2v, 1-\alpha/2} ,$$

$$\frac{2v}{\chi_{2v, w}^2} \approx F_{n, 2v, 1-w} ,$$

donde $\chi_{2v, w}^2$ es el cuartil w de una distribución *chi-cuadrado* con $2v$ grados de libertad. Ahora, sustituyendo en nuestra fórmula y despejando n , obtenemos

$$n \geq \frac{8\rho z_{1-\alpha/2}}{l^2 \chi_{2v, w}^2} . \quad (24)$$

4.3.3. Comparación de criterios

En esta sección vamos a comparar los tamaños muestrales obtenidos a partir de cada uno de los criterios cuando la precisión τ es desconocida. Para ello, vamos a tomar como referencia las comparaciones realizadas en [5].

En primer lugar, vamos a comparar los tamaños muestrales obtenidos con ACC y ALC. Para ello, vamos a suponer que n es suficientemente grande, después despejamos $n + n_0$ y realizamos el cociente entre las fórmulas obtenidas con cada método.

Tomamos la ecuación (19) obtenida a partir del criterio ACC y la aproximación obtenida a partir de ALC (21). Cuando $n \rightarrow \infty$, obtenemos

$$\frac{(n + n_0)_{ACC}}{(n + n_0)_{ALC}} = \frac{t_{2v, 1-\alpha/2}^2}{z_{1-\alpha/2}^2} C_v , \quad (25)$$

donde C_v viene dada por la expresión (22), y $(n + n_0)_{ACC}$, $(n + n_0)_{ALC}$ son los tamaños muestrales totales obtenidos a partir de ACC y ALC respectivamente.

Ahora, para cada $v > 0,5$, existe un α_v que nos permite diferenciar los casos en los que esta expresión es mayor o menor que 1, con v el primer parámetro de la distribución gamma que sigue la precisión τ .

Obtenemos

$$\frac{t_{2v,1-\alpha/2}^2}{z_{1-\alpha/2}^2} C_v \begin{cases} > 1 & \text{si } \alpha < \alpha_v \\ = 1 & \text{si } \alpha = \alpha_v \\ < 1 & \text{si } \alpha > \alpha_v \end{cases}.$$

Por tanto, comparando el tamaño muestral total $n + n_0$ de cada método, podemos observar

$$(n + n_0)_{ACC} \begin{cases} > (n + n_0)_{ALC} & \text{si } 1 - \alpha > 1 - \alpha_v \\ = (n + n_0)_{ALC} & \text{si } 1 - \alpha = 1 - \alpha_v \\ < (n + n_0)_{ALC} & \text{si } 1 - \alpha < 1 - \alpha_v \end{cases}.$$

Tenemos que α_v es una función que depende únicamente de v , y es creciente respecto de ella. Ahora, podemos observar que las diferencias entre los tamaños muestrales obtenidos por cada método dependen únicamente de los valores de v y α . Es decir, dado v , cuando la distancia entre α_v y α aumenta, entonces la diferencia entre los tamaños muestrales aumenta, y α_v es el punto en el que los tamaños muestrales obtenidos por ambos métodos coinciden.

Ejemplo 4.4 *Vamos a rescatar el estudio clínico visto en el ejemplo 3.1.3 sobre el efecto del consumo de leche agria en la hipertensión. Anteriormente, suponíamos que $\sigma^2 = 25$; es decir, $\tau = 1/25$. Ahora, vamos a suponer que τ es desconocida, con $\tau \sim \Gamma(v, \rho)$. La idea principal para escoger estos parámetros y poder comparar los nuevos resultados con los obtenidos a lo largo de la Sección 3, será asegurar que los parámetros de esta distribución cumplan que la esperanza de τ ; dada por $E(\tau) = v/\rho$, sea igual a $1/25$. Como ejemplo, vamos a tomar los valores $v = 2$ y $\rho = 50$, propuestos en [5]. Además, vamos a tomar $n_0 = 10$. En la Figuras 4 y la Figura 5 podemos observar la evolución de los tamaños muestrales para los criterios ACC y ALC, con $l = 1$ y $l = 2$, en función del nivel de probabilidad de cobertura del parámetro $1 - \alpha$. Observamos que tomando un valor $l = 1$, $\alpha_v = 0,867$; y, tomando un valor $l = 2$, $\alpha_v = 0,863$.*

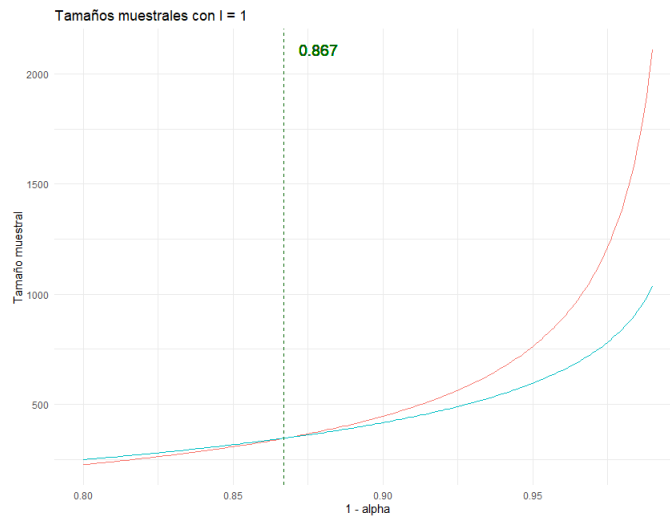


Figura 4: Comparación de criterios ACC y ALC ($l = 1$, $v = 2$, $\rho = 50$ y $n_0 = 10$).

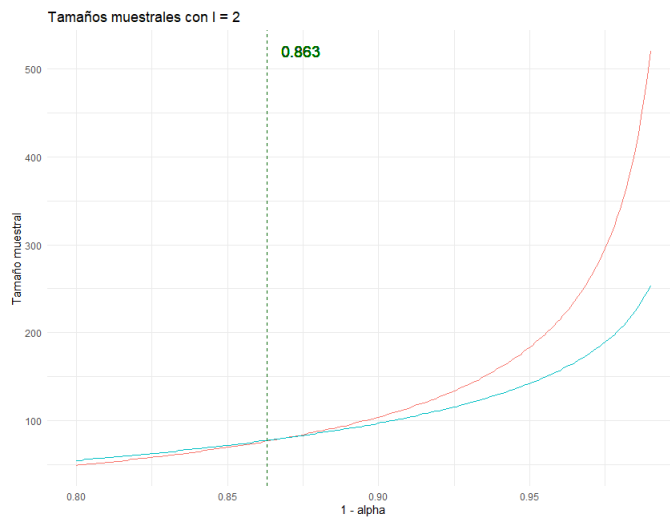


Figura 5: Comparación de criterios ACC y ALC ($l = 2$, $v = 2$, $\rho = 50$ y $n_0 = 10$).

Por otro lado, también es interesante analizar la evolución del punto de corte α_v en función de distintos valores de v , con el fin de saber dónde un método devolverá un tamaño muestral mayor que el otro. En este caso, vamos a tomar los valores $\rho = 50$, $n_0 = 10$ y $l = 1$.

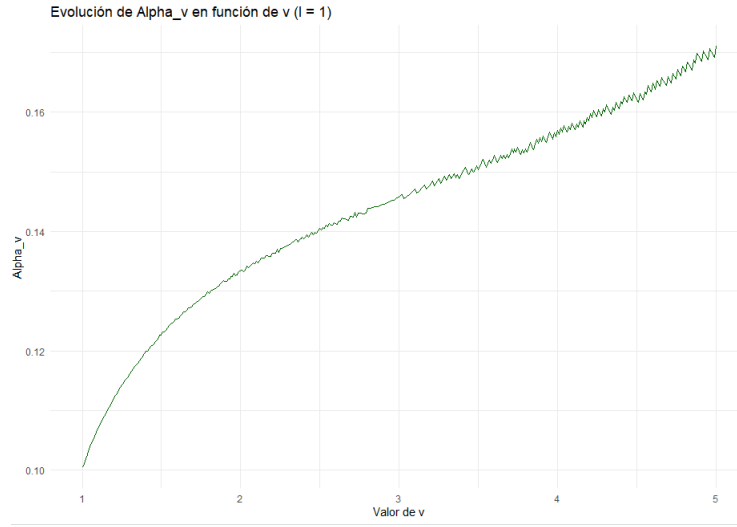


Figura 6: Evolución del valor de α_v respecto de v ($l = 1$, $\rho = 50$ y $n_0 = 10$).

En la figura 6, podemos observar cómo evoluciona este punto de corte α_v en función de distintos valores de v . Además, podemos deducir, que en este ejemplo concreto, para cualquier $\alpha < 0,10$, el tamaño muestral propuesto obtenido con el método ACC siempre va a ser mayor que el obtenido con ALC, esto se debe a que el método ALC es más sensible al nivel de probabilidad $1 - \alpha$, ya que es el valor fijado a la hora de trabajar.

En segundo lugar, vamos a comparar el tamaño muestral obtenido con el criterio WOC y el ACC. Realizando el cociente entre las aproximaciones (20) y (24), obtenemos

$$\frac{(n + n_0)_{WOC}}{(n + n_0)_{ALC}} = \frac{2v}{\chi_{2v,w}^2} C_v. \quad (26)$$

Dado $v > 0,5$ y un valor de w pequeño (normalmente $w = 0,05$), el ratio obtenido en la expresión (26) será siempre mayor que 1.

Finalmente, comparamos WOC y ACC. Realizando el cociente entre las expresiones (19) y (24), obtenemos

$$\frac{(n + n_0)_{WOC}}{(n + n_0)_{ACC}} = \frac{2v z_{1-\alpha/2}^2}{\chi_{2v,w}^2 t_{2v,1-\alpha/2}^2}. \quad (27)$$

Del mismo modo que antes, este ratio será siempre mayor que 1.

Ejemplo 4.5 La Figura 7 y la Figura 8 muestran los distintos tamaños muestrales de los tres métodos bayesianos para los datos del Ejemplo 4.4. Podemos ver cómo con el criterio WOC siempre logramos un tamaño muestral superior al resto, y además, vemos como a

medida que vamos disminuyendo el valor de w , obtenemos valores más pequeños, ya que el criterio será menos restrictivo.

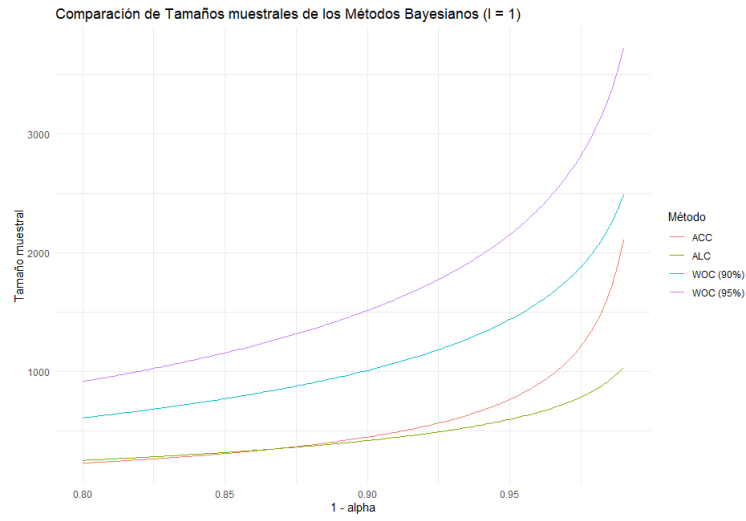


Figura 7: Comparación de criterios ACC, ALC y WOC ($l = 1$, $v = 2$, $\rho = 50$ y $n_0 = 10$).

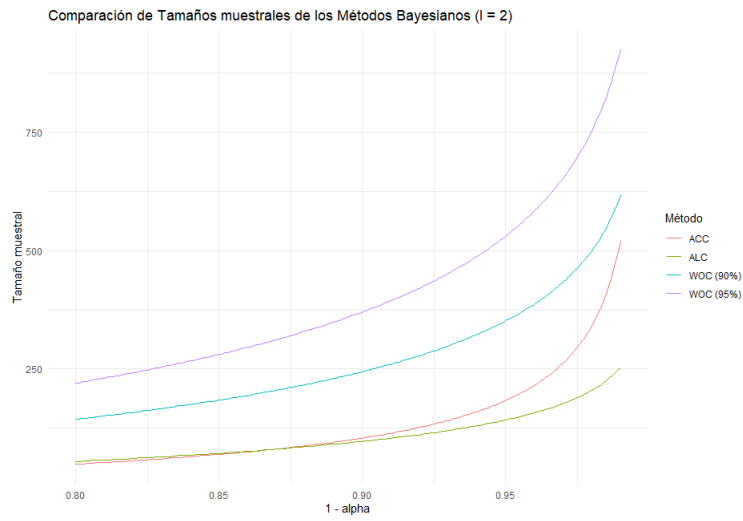


Figura 8: Comparación de criterios ACC, ALC y WOC ($l = 2$, $v = 2$, $\rho = 50$ y $n_0 = 10$).

4.4. Estimación de una diferencia de medias. Criterios y comparación

En esta sección, vamos a aplicar la idea de los criterios de ACC, ALC y WOC a una diferencia de medias para los experimentos de dos grupos con el diseño paralelo con misma precisión (misma varianza). Supongamos que tenemos dos muestras $x_1 = (x_{11}, \dots, x_{n_11})$ y $x_2 = (x_{12}, \dots, x_{n_22})$ obtenidas a partir de dos tratamientos diferentes. Normalmente, tendremos una muestra con gente con la que hemos aplicado un tratamiento ya conocido y otra muestra con gente que prueba el nuevo tratamiento. Supongamos que $x_{ij}, i = 1, \dots, n_j$ son independientes y están idénticamente distribuidas con una distribución normal con media μ_j y precisión común τ . De manera similar al problema de la Sección 4.3, suponemos que, a priori, las variables μ_j y τ siguen las distribuciones

$$\tau \sim \Gamma(v, \rho) ,$$

$$\mu_j | \tau \sim N(\mu_{0j}, n_{0j}\tau) , j = 1, 2 .$$

En esta ocasión, la idea para poder aplicar los criterios anteriores será tomar $\theta = \mu_1 - \mu_2$ como parámetro de interés sobre el que en un estudio posterior se buscará realizar una estimación.

4.4.1. Precisión conocida

Supongamos que la precisión τ es conocida, procediendo de una manera similar al apartado anterior, obtenemos que la diferencia entre las medias θ sigue la siguiente distribución:

$$\theta | x_1, x_2 \sim N \left(\mu_{n_22} - \mu_{n_11}, \frac{\tau(n_{01} + n_1)(n_{02} + n_2)}{n_1 + n_2 + n_{01} + n_{02}} \right) ,$$

donde

$$\mu_{n_jj} = \frac{n_{0j}\mu_{0j} + n_j\bar{x}_j}{n_{0j} + n_j} , j = 1, 2 .$$

Podemos observar que la varianza posterior de θ depende exclusivamente de los tamaños muestrales n_1 y n_2 . Estamos en una situación parecida a la de la sección anterior, nos encontramos con que los tres métodos nos proporcionan el mismo resultado.

Supongamos que buscamos un tamaño muestral igual para cada uno de los dos grupos, entonces, tomando como referencia el desarrollo propuesto en [9], obtenemos:

$$n_1 = n_2 \geq \frac{-B + \sqrt{B^2 - 4AC}}{2A} , \quad (28)$$

donde

$$\begin{aligned} A &= \tau^2 , \\ B &= \tau^2(n_{01} + n_{02}) - 28 \frac{z_{1-\alpha/2}^2}{l^2} , \\ C &= \tau^2 n_{01} n_{02} - \frac{4(n_{01} + n_{02}) \tau z_{1-\alpha/2}^2}{l^2} . \end{aligned}$$

Observación 4.6 *En la práctica, si obtenemos $B^2 - 4AC < 0$, significa que el tamaño muestral previo (n_{01}, n_{02}) ya es suficiente y no necesitamos una muestra mayor.*

Observando la distribución de θ , podemos minimizar la varianza posterior haciendo $n_1 + n_{01} = n_2 + n_{02}$. En esta situación, el menor tamaño muestral necesario, puede ser obtenido resolviendo las ecuaciones

$$\begin{aligned} n_1 &\geq \frac{8}{\tau l^2} z_{1-\alpha/2}^2 - n_{01} , \\ n_2 &= n_1 + n_{01} - n_{02} . \end{aligned}$$

Del mismo modo que en la Sección 4.3.1, cuando la varianza es conocida, los tamaños muestrales obtenidos a partir de estas ecuaciones son bastante similares a los propuestos por los autores frecuentistas. De hecho, si tomamos igual que antes $n_1 = kn_2$ y tomamos el sistema de ecuaciones (3), podemos observar que las fórmulas bayesianas nunca devuelven un tamaño muestral superior; y, si además tomamos $n_{01} = n_{02} = 0$, las ecuaciones obtenidas aportan los mismos resultados.

4.4.2. Precisión desconocida

Supongamos ahora que no conocemos la precisión común τ . Entonces, tomando los resultados de [6], tenemos que nuestro parámetro de interés θ sigue la siguiente distribución:

$$\theta|x_1, x_2 \sim A + t_{2v} \sqrt{\frac{B}{2CD}} ,$$

donde

$$\begin{aligned} A &= E(\theta|x_1, x_2) = \frac{n_1 \bar{x}_1 + n_{01} \mu_{01}}{n_1 + n_{01}} = \frac{n_2 \bar{x}_2 + n_{02} \mu_{02}}{n_2 + n_{02}} , \\ B &= 2\rho + n_1 s_1^2 + n_2 s_2^2 + \frac{n_1 n_{01}}{n_1 + n_{01}} (\bar{x}_1 - \mu_{01})^2 + \frac{n_2 n_{02}}{n_2 + n_{02}} (\bar{x}_2 - \mu_{02})^2 , \\ C &= \frac{n_1 + n_2}{2} + v , \\ D &= \frac{(n_1 + n_{01})(n_2 + n_{02})}{n_1 + n_{01} + n_2 + n_{02}} . \end{aligned}$$

Average Coverage Criterion (ACC)

Procediendo de una manera similar a la del apartado anterior, basándonos en [9], llegamos a la conclusión de que buscaremos los menores n_j que cumplan la expresión

$$\frac{(n_1 + n_{01})(n_2 + n_{02})}{n_1 + n_{01} + n_2 + n_{02}} \geq \frac{4\rho}{vl^2} t_{2v,1-\alpha/2}^2 . \quad (29)$$

Ahora, igual que en la sección anterior, si trabajamos con la estrategia de obtener para cada grupo el mismo tamaño muestral, es decir, $n_1 = n_2$, podemos obtener el tamaño muestral resolviendo la siguiente inecuación

$$n_1 = n_2 \geq \frac{-B + \sqrt{B^2 - 4AC}}{2A} , \quad (30)$$

donde

$$A = \frac{vl^2}{4} ,$$

$$B = \frac{vl^2}{4}(n_{01} + n_{02}) - 2\rho t_{2v,1-\alpha/2}^2 ,$$

$$C = \frac{n_{01}n_{02}vl^2}{4} - \rho t_{2v,1-\alpha/2}^2(n_{01} + n_{02}) .$$

Average Length Criterion (ALC)

De manera similar, aplicando el criterio ALC obtenemos la expresión

$$2t_{n_1+n_2+2v,1-\alpha/2} \sqrt{\frac{2\rho(n_1 + n_{01} + n_2 + n_{02})}{(n_1 + n_2 + 2v)(n_1 + n_{01})(n_2 + n_{02})}} \frac{\Gamma\left(\frac{n_1 + n_2 + 2v}{2}\right) \Gamma\left(\frac{2v-1}{2}\right)}{\Gamma\left(\frac{n_1 + n_2 + 2v-1}{2}\right) \Gamma\left(\frac{2v}{2}\right)} \leq l . \quad (31)$$

Ahora, aplicando la misma técnica de aproximación que en el caso de (20), basándonos en las aproximaciones propuestas en [5], cuando el valor $n_1 + n_2$ es suficientemente grande, podemos aproximar esta ecuación para obtener:

$$\frac{(n_1 + n_{01})(n_2 + n_{02})}{n_1 + n_{01} + n_2 + n_{02}} \leq \frac{4\rho z_{1-\alpha/2}^2}{l^2 v C_v} , \quad (32)$$

donde C_v es la obtenida en la expresión (22).

WOC

Procediendo igual que anteriormente, aplicando el criterio WOC tomando como referencia [9], llegamos a la conclusión de que se buscarán los menores n_1, n_2 que cumplan la inecuación

$$\frac{l^2(n_1 + n_{01})(n_2 + n_{02})(n_1 + n_2 + 2v)}{8\rho(n_1 + n_{01} + n_2 + n_{02}) \left(1 + \frac{n_1 + n_2}{2v} F_{n_1+n_2, 2v, 1-w}\right)} \geq t_{n_1+n_2+2v, 1-\alpha/2}^2. \quad (33)$$

Ahora, de la misma manera que en la sección anterior, basándonos en los cálculos realizados en [5], cuando $n_1 + n_2$ es suficientemente grande, podemos aproximar esta ecuación para trabajar con ella de manera más sencilla. Obtenemos:

$$\frac{(n_1 + n_{01})(n_2 + n_{02})}{n_1 + n_{01} + n_2 + n_{02}} \leq \frac{8\rho z_{1-\alpha/2}^2}{l^2 v C_v}, \quad (34)$$

con C_v es la obtenida en la expresión (22).

4.4.3. Comparación de criterios

Igual que en el apartado de una muestra, vamos a basarnos en el artículo [5] para realizar estas comparaciones.

Por un lado, cuando la varianza σ^2 es conocida, hemos visto que la fórmula obtenida por los tres criterios es la misma, por lo tanto no hay nada que comparar.

Supongamos entonces que la precisión τ es desconocida. Tenemos que cada criterio nos proporciona ecuaciones diferentes. Vamos a suponer que $n_1 + n_{01} = n_2 + n_{02}$, y vamos a dejar entonces las expresiones en función de $n_1 + n_{01}$.

En primer lugar, vamos a comparar ACC y ALC. Dividiendo las fórmulas y operando, obtenemos:

$$\frac{(n_1 + n_{01})_{ACC}}{(n_1 + n_{01})_{ALC}} = \frac{t_{2v, 1-\alpha/2}^2}{z_{1-\alpha/2}^2} C_v. \quad (35)$$

En segundo lugar, vamos a comparar WOC y ALC. Obtenemos

$$\frac{(n_1 + n_{01})_{WOC}}{(n_1 + n_{01})_{ALC}} = \frac{2v}{\chi_{2v, w}^2} C_v. \quad (36)$$

Finalmente, comparamos los criterios WOC y ACC, obtenemos:

$$\frac{(n_1 + n_{01})_{WOC}}{(n_1 + n_{01})_{ACC}} = \frac{2v z_{1-\alpha/2}^2}{\chi_{2v, w}^2 t_{2v, 1-\alpha/2}^2}. \quad (37)$$

Si nos fijamos en los resultados obtenidos, tenemos exactamente las mismas aproximaciones que en la Sección 4.3.3, cuando estudiábamos la estimación de una media, por lo que los análisis y las conclusiones vistas anteriormente pueden extrapolarse a las fórmulas para la estimación de una diferencia de medias.

5. Comparación entre métodos bayesianos y frecuentistas

A lo largo de esta sección vamos a centrarnos en los métodos vistos para la estimación de una media. Para realizar una comparación más visual, he elaborado un programa con el paquete *Shiny* en *R*, donde introducimos los parámetros de nuestro problema y observamos la evolución de los distintos tamaños muestrales para cada uno de los métodos. En el Anexo B se puede consultar el código y una imagen general del programa.

En primer lugar, vamos a comparar los métodos aplicando los datos del Ejemplo 3.1.3. En la Figura 10 podemos observar una tabla con los resultados de los distintos tamaños muestrales para $\alpha = 0,05$ y $l = 1$; y, en la Figura 11 podemos observar la evolución de los tamaños muestrales en función del valor escogido $1 - \alpha$. Cabe destacar que para trabajar con los métodos en los que la varianza σ^2 es conocida, el programa toma por defecto $\sigma^2 = \rho/v$, ya que será la esperanza del parámetro. Como hemos visto en la Observación 4.2, con los tres métodos bayesianos obtenemos el mismo resultado y este tamaño muestral nunca es superior al obtenido mediante la fórmula frecuentista.

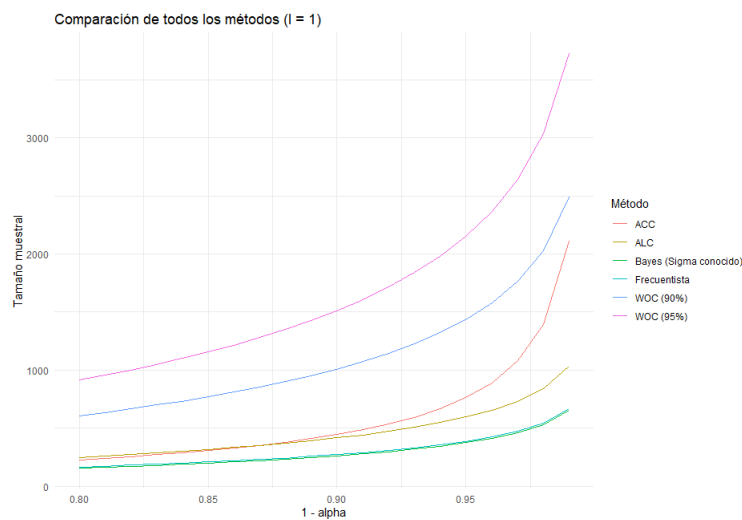


Figura 9: Gráfica con una comparación de todos los criterios para la estimación de una media ($l = 1$, $v = 2$, $\rho = 50$ y $n_0 = 10$).

Tamaño Muestral ($l = 1$, Alpha: 0.05)

Criterio	n
1 Frecuentista	385.00
2 Bayes (sigma conocido)	375.00
3 ACC	761.00
4 ALC	595.00
5 WOC(90%)	1435.00
6 WOC(95%)	2152.00

Figura 10: Tabla con una comparación de todos los criterios para la estimación de una media para $\alpha = 0,05$ ($l = 1$, $v = 2$, $\rho = 50$ y $n_0 = 10$).

En este ejemplo concreto, los tamaños muestrales obtenidos con las fórmulas para varianza conocida, siempre son menores que el resto, cosa que tiene sentido ya que contamos con una información clave sobre nuestros datos. Sin embargo, esto no tiene por qué ser siempre así, debido a que la fórmula (1) puede ser bastante sensible a los valores de σ^2 y l . En la Figura 11 podemos observar un ejemplo en el que el resultado obtenido a través de los métodos bayesianos ACC y ALC es menor.

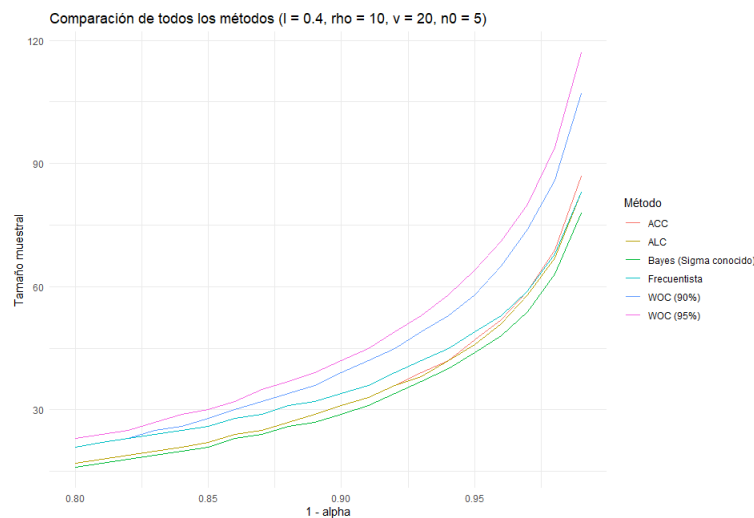


Figura 11: Gráfica con una comparación de todos los criterios para la estimación de una media ($l = 0,04$, $v = 20$, $\rho = 10$ y $n_0 = 5$).

6. Conclusiones

A lo largo del trabajo hemos analizado distintos métodos con el fin de hallar el tamaño muestral necesario para poder obtener unas buenas conclusiones a la hora de realizar un estudio. Hemos comenzado viendo los métodos propuestos por los investigadores de la corriente frecuentista, tanto para la estimación como para la comparación de medias, aplicando intervalos de confianza y contrastes de hipótesis; y hemos finalizado estudiado tres criterios diferentes basados en la estadística bayesiana para la estimación de una media. La teoría de los distintos métodos se ha visto complementada con ejemplos de estudios relacionados con la investigación clínica.

Los principales objetivos del trabajo eran: por un lado, tener una visión clara de cómo funcionan las fórmulas clásicas utilizadas en investigación clínica, es decir, las propuestas por los autores frecuentistas; y, por otro lado, realizar una comparativa con los métodos bayesianos, debido a que, a la hora de buscar el tamaño muestral, estamos en un estado previo a la obtención de datos y necesitamos realizar suposiciones sobre el comportamiento de nuestra futura muestra basándonos en estudios previos o información subjetiva, por lo que pensamos que era interesante analizarlo desde el punto de vista de la estadística bayesiana, ya a que estas suposiciones estarían recogidas de manera más transparente en la llamada distribución a priori.

Hemos analizado los métodos más estudiados, suponiendo muestras normales y distinguiendo los casos en los que el valor de σ^2 es conocido o desconocido. Sin embargo, a la hora de realizar un estudio, normalmente esta varianza será desconocida. Como solución, los autores frecuentistas recomiendan realizar una estimación de este parámetro y trabajar directamente como si fuera conocida a priori. Observamos que, en estos casos en los que suponemos que la varianza es conocida, los métodos bayesianos nunca aportan un tamaño muestral superior al obtenido con los métodos frecuentistas.

Además, en los casos en los que trabajamos suponiendo que la varianza es desconocida, la corriente bayesiana aporta una solución mucho más lógica y natural que los otros métodos. Debido a que se trabaja a partir de una deducción de la distribución que seguirá este parámetro, recogida en la distribución a priori, en lugar de realizar directamente una estimación del mismo, cosa que es más sencilla de intuir a priori y tiene más margen de error. También es cierto que, a la hora de realizar ciertos ensayos, si tenemos mucha información sobre la varianza, puede ser que, al suponerla desconocida y trabajar con estos métodos bayesianos, obtengamos un número de individuos excesivamente alto.

En conclusión, hemos visto cómo tratar el tema del tamaño muestral desde una pers-

pectiva bayesiana es una buena alternativa a las clásicas fórmulas frecuentistas, debido a que aporta soluciones más flexibles, especialmente cuando los datos son escasos y no contamos con mucha información sobre la población a estudiar. Pero, por otro lado, siempre hay que tener presente la idea de analizar cada ensayo de manera independiente y particular, estudiando la información previa con la que se cuenta sobre la muestra, para así poder concluir si merece la pena utilizar unos métodos u otros.

Referencias

- [1] Adcock, C.J. (1988) A Bayesian Approach to Calculating Sample Sizes. *Journal of the Royal Statistical Society. Series D (The Statistician)*. 37(4/5): 433-439
- [2] Adcock, C.J. (1997) Sample Size Determination: A Review. *Journal of the Royal Statistical Society. Series D (The Statistician)*. 46(2): 261-283.
- [3] Bernardo, J.M. and Smith, A.F.M. (1998) *Bayesian Theory*. John Wiley & Sons.
- [4] Box, G.E. and Tiao, G.C. (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley.
- [5] Cao, J., Lee, J.J., and Albert, S. (2009) Comparison of Bayesian sample size criteria: ACC, ALC, and WOC. *J Stat Plan Inference*. 139(12): 4111-4122.
- [6] Chow, S.C., Shao, J., Wang, H., y Lokhnygina, Y. (2020). *Sample Size Calculations in Clinical Research*. CRC Press.
- [7] Correa Morales, J.C. and Barrera Causil, C.J. (2019) *Introducción a La Estadística Bayesiana*. Instituto Tecnológico Metropolitano.
- [8] Herrera, R., et al. (2012) Modelos de la Familia Exponencial. *Ingeniare*, 12: 89-98
- [9] Joseph, L., y Belisle, P. (1997) Bayesian sample size determination for normal means and differences between normal means. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(2): 209-226.
- [10] Joseph, L., Wolfson, D.B., y Du Berger, R. (1995) Sample Size Calculations for Binomial Proportions via Highest Posterior Density Intervals. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 44(2): 143-154.
- [11] Julious, S.A. (2004) Sample sizes for clinical trials with Normal data. *Statistics in Medicine*, 23(12): 1921-1986.
- [12] Kieser, M. (2020) *Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials*. Springer.
- [13] Machin, D., Campbell, M.J., Tan, S.B., y Tan, S.H. (2008) *Sample Size Tables for Clinical Studies*. Wiley-Blackwell

-
- [14] Mizushima, S., Ohshige, K., Watanabe, J., Kimura, M., Kadowaki, T., Nakamura, Y., Tochikubo, O., y Ueshima, H. (2004) Randomized controlled trial of sour milk on blood pressure in borderline hypertensive men. *American Journal of Hypertension*, 17(8): 701-706.

A. Detalles del desarrollo del trabajo

La Tabla 3 recoge la información respectiva al tiempo aproximado dedicado para cada uno de los puntos relevantes de la elaboración del trabajo. Y, la Tabla 4 recoge las asignaturas cursadas a lo largo del grado con las que el trabajo está relacionado.

Tarea	Tiempo (horas)
Recopilación de materiales	30
Estudio de bibliografía	40
Elaboración de resultados gráficos/numéricos	30
Redacción de la memoria	50
Total	150

Tabla 3: *Tiempo aproximado de dedicación al trabajo*

Asignatura	Páginas	Descripción
Inferencia estadística	8-11	La sección 2.1 está elaborada a partir de los apuntes.
Inferencia estadística	22, 25	Los desarrollos iniciales de las secciones 3.2.1.1 y 3.2.1.2 se obtienen a partir de los apuntes.
Probabilidad	General	Hay relación, pero no es referencia básica.
Introducción a la estadística	General	Hay relación, pero no es referencia básica.

Tabla 4: *Asignaturas relacionadas con el trabajo*

B. Código de R

Aquí tenemos el código utilizado para crear el programa *Shiny* con R. Hemos utilizado el paquete *SampleSizeMeans*, el cual contiene los distintos métodos estudiados para obtener el tamaño muestral adecuado en ensayos cuyo fin es la estimación de una media.

```
1 library("shiny")
2 library("SampleSizeMeans")
3 library(ggplot2)
4 library(shinythemes)
5
6 # Definición del UI
7 ui <- fluidPage(
8   theme = shinytheme("flatly"), # Aplicar un tema
9   titlePanel("Cálculo del Tamaño Muestral para estimación de una
10     media"),
11
12   sidebarLayout(
13     sidebarPanel(
14       numericInput("v", "v:", 2, min = 0),
15       numericInput("rho", "rho:", 2),
16       numericInput("long", "longitud del intervalo:", 0.2),
17       numericInput("n0", "n0:", 10),
18       sliderInput("alpha", "Alpha:", min = 0, max = 0.2, value = 0.05,
19         step = 0.01),
20       actionButton("calculate", "Calcular")
21     ),
22     mainPanel(
23       fluidRow(
24         column(12, uiOutput("table_with_title"))
25       ),
26       fluidRow(
27         column(12, plotOutput("sample_size_plot"))
28       )
29     )
30   )
31
32 # Definición del servidor
33 server <- function(input, output) {
34   observeEvent(input$calculate, {
35     v <- input$v
36     rho <- input$rho
```

```

36 len <- input$long
37 n0 <- input$n0
38 alpha <- input$alpha
39 lambda <- v / beta
40
41 results <- data.frame(
42   Criterio = c("Frecuentista", "Bayes (sigma conocido)", "ACC",
43     "ALC", "WOC(90%)", "WOC(95%)"),
44   n = c(mu.freq(len, lambda, 1- alpha),
45     mu.freq(len, lambda, 1- alpha) - n0,
46     mu.acc(len, v, rho, n0, 1- alpha),
47     mu.alc(len, v, rho, n0, 1- alpha),
48     mu.modwoc(len, v, rho, n0, 1- alpha, 0.9),
49     mu.modwoc(len, v, rho, n0, 1- alpha, 0.95))
50 )
51
52 output$result_table <- renderTable({
53   results
54 }, rownames = TRUE) # Mostrar nombres de filas si es necesario
55
56 output$table_with_title <- renderUI({
57   tagList(
58     h4(paste("Tamaño Muestral ( Alpha:", round(alpha, 2), ")")),
59     tableOutput("result_table")
60   )
61 })
62
63 levels <- seq(0.8, 0.95, by = 0.01)
64 mu_acc_values <- sapply(levels, function(level) mu.acc(len, v, rho,
65   n0, level))
66 mu_alc_values <- sapply(levels, function(level) mu.alc(len, v, rho,
67   n0, level))
68 mu_woc_90_values <- sapply(levels, function(level) mu.modwoc(len,
69   v, rho, n0, level, 0.90))
70 mu_woc_95_values <- sapply(levels, function(level) mu.modwoc(len,
71   v, rho, n0, level, 0.95))
72 mu_freq_values <- sapply(levels, function(level) mu.freq(len,
73   lambda, level))
74 mu_bay_values <- sapply(levels, function(level) mu.freq(len,
75   lambda, level)-n0)
76
77 plot_data <- data.frame(
78   level = levels,

```



```
72     mu_acc = mu_acc_values,
73     mu_alc = mu_alc_values,
74     mu_woc_90 = mu_woc_90_values,
75     mu_woc_95 = mu_woc_95_values,
76     mu_freq = mu_freq_values,
77     mu_bay = mu_bay_values
78   )
79
80   output$sample_size_plot <- renderPlot({
81     ggplot(plot_data, aes(x = level)) +
82       geom_line(aes(y = mu_acc, color = "ACC")) +
83       geom_line(aes(y = mu_alc, color = "ALC")) +
84       geom_line(aes(y = mu_woc_90, color = "WOC (0.90)")) +
85       geom_line(aes(y = mu_woc_95, color = "WOC (0.95)")) +
86       geom_line(aes(y = mu_freq, color = "Frecuentista")) +
87       geom_line(aes(y = mu_freq, color = "Bayes (sigma conocido)")) +
88       labs(title = "Comparación de todos los métodos",
89            x = "1 - alpha",
90            y = "Tamaño muestral",
91            color = "Método") +
92       theme_minimal()
93   })
94 })
95 }
96
97 # Ejecución de la aplicación Shiny
98 shinyApp(ui = ui, server = server)
```

Listing 1: Código del programa creado con Shiny en R.

Finalmente, en la Figura 12 podemos observar cómo se vería el programa al ejecutarlo. La idea es poder seleccionar los valores adecuados para los distintos parámetros, y dibujar una tabla con el tamaño muestral que proporciona cada uno de los métodos y una gráfica con la evolución en función del valor de α .

Cálculo del Tamaño Muestral para estimación de una media

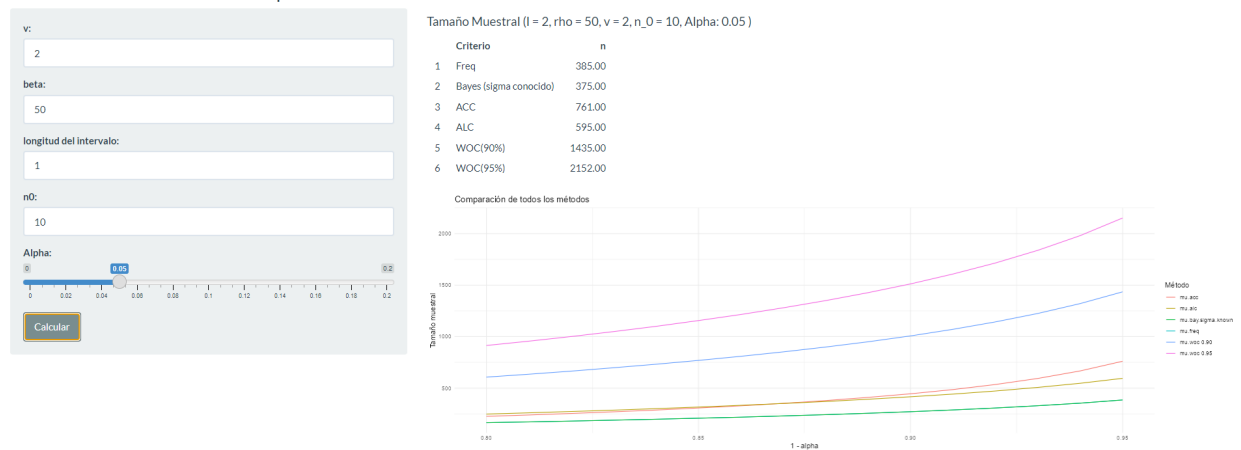


Figura 12: Interfaz del programa creado para realizar comparaciones de los tamaños muestrales obtenidos para la estimación de una media con cada uno de los métodos.