

37 Congreso
Nacional
CENTRO DE
CONVENCIONES
INTERNACIONALES

Barcelona
22/25
MAYO 2024

seram
Sociedad Española de Radiología Médica

FERM
FUNDACIÓN ESPAÑOLA DE RADIOLOGÍA MÉDICA

RC | RADIOLEGS
DE CATALUNYA

¿Cómo saco adelante la parte estadística de mi trabajo? - Una nueva perspectiva.

Nader Salhab Ibáñez

Hospital Clínic Universitari de València, València

Objetivo docente

Introducción

La estadística nos rodea. Tanto en nuestra vida privada como en nuestra vida profesional nuestro cerebro ha evolucionado para detectar patrones, correlaciones y realizar inferencias. La estadística muchas veces nos puede producir rechazo o ser una de las partes más tediosas a las que nos tenemos que enfrentar cuando llevamos a cabo un estudio. Quizá tenga que ver con su componente matemático o qué, en la carrera de medicina – al menos en España –, se suele ver en una asignatura aislada que se da al principio de la carrera muy desconectada del resto de asignaturas.

Este trabajo busca conseguir cambiar la mentalidad del lector respecto a la estadística. Mostrar que, al menos la parte más básica que es la que nos concierne – no entrando en la parte matemática que sí presenta una alta complejidad, especialmente para quién no ha recibido una formación matemática adecuada – no es tan abrumadora como podemos pensar. Además, se mostrará cómo funciona y cómo nos puede ayudar a entender mejor el mundo de la estadística la llamada estadística probabilística o Bayesiana en sustitución de la estadística frecuentista que es a la que estamos acostumbrados.

En este trabajo, a parte de la parte más teórica, habrá una parte práctica donde se mostrará cómo funciona la estadística probabilística en contraposición a la frecuentista. Igualmente, se utilizará para todos los ejemplos el lenguaje R con el entorno de desarrollo integrado RStudio en sus versiones más actualizadas al momento de la realización de los ejemplos (información que será dada más adelante) – en todos los ejemplos de esta presentación las versiones son las mismas.

Además, se recomienda tanto el lenguaje R como RStudio como programas de estadística de uso diario tanto por su versatilidad, por ser de código abierto, por ser gratuitos y por ser ofrecer un amplio abanico de herramientas estadísticas.

37 Congreso
Nacional
CENTRO DE
CONVENCIONES
INTERNACIONALES

Barcelona
22/25
MAYO 2024

seram
Sociedad Española de Radiología Médica

FERM
FUNDACIÓN ESPAÑOLA DE RADIOLOGÍA MÉDICA

RC | RADIOLEGS
DE CATALUNYA

Revisión del tema

Estadística frecuentista. ¿De dónde venimos?

Puede que el nombre “estadística frecuentista” no nos diga nada a primeras, pero, simplificando, sería la estadística – el conjunto de herramientas y conceptos con los que trabajamos – que utilizamos habitualmente. ¿Cuándo desarrollas la parte estadística utilizas el “p valor”?, ¿intervalos de confianza?, ¿contrastos de hipótesis?... Entonces, estás utilizando estadística frecuentista.

Esta interpretación de la probabilidad se basa en las frecuencias. Relaciona las frecuencias de un evento con la probabilidad de ocurrencia de dicho evento. Por ejemplo, si tenemos un dado estándar de 6 caras y lo lanzamos un total de 1000 veces, cada cara habrá salido aproximadamente $1/6$ de todas las tiradas. A mayor número de tiradas que hagamos más nos acercaremos a la probabilidad real teórica de $1/6$ – más cerca de lo que se llama *límite* de su frecuencia relativa estaremos – y por tanto podremos inferir que la probabilidad de que salga cada una de las caras de un dado es $1/6$ (**Figura 1**).

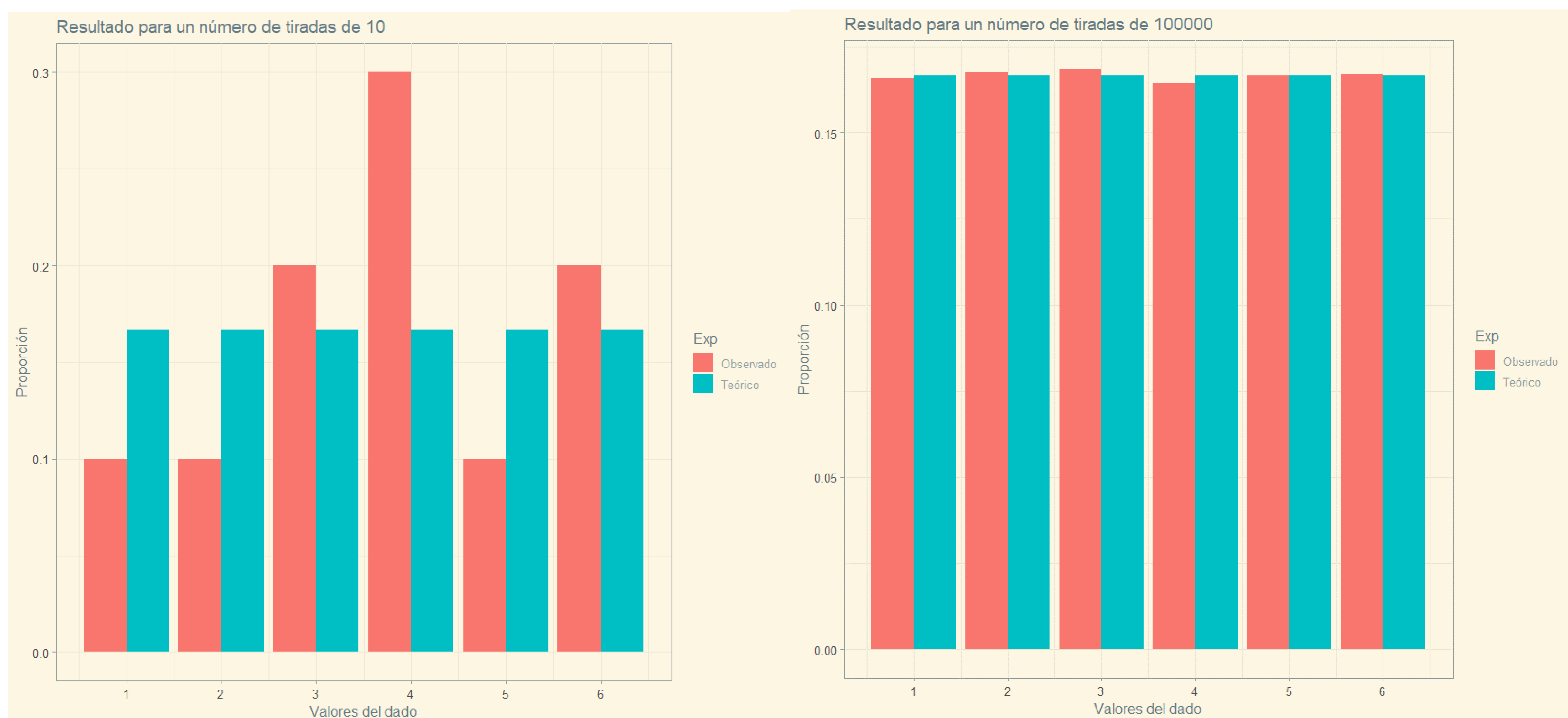


Figura 1. Resultado de una tirada de dados, en la gráfica de la derecha de un total de 10 y en la de la izquierda de 100000. Se puede ver que, a mayor número de tiradas, más nos acercamos al valor teórico real de la proporción con la que sale cada cara.

Este tipo de estadística es la más utilizada hoy en día, especialmente el *Null Hypothesis Significance Testing* (NHST) – si aceptamos o no la hipótesis nula.

Cuando realizamos un estudio para valorar si un determinado fármaco, o un determinado tipo de prueba de imagen, es mejor que otro (entendiendo como mejor que cure un mayor número de infecciones o que diagnostique un mayor número de patología, por ejemplo), buscamos que tipo de herramienta estadística debemos utilizar para realizar este tipo de inferencia, ¿una t de Student?, ¿o quizá una prueba robusta ya que nuestras variables no han superado el test de normalidad?, ¿y si quiero comparar más de dos grupos? En todos estos casos, nos vamos a encontrar con pruebas que utilizan la inferencia frecuentista para responder a nuestras preguntas. También encontraremos conceptos como el Intervalo de Confianza (IC) o el p valor que, de nuevo, pertenecen al ámbito de la interpretación frecuentista de la estadística.

No hay duda, entonces, de que **la estadística frecuentista es la más utilizada**. Estamos acostumbrados a hablar de ella, de p valor, de intervalos de confianzas, de pruebas como ANOVA... Mientras que la estadística Bayesiana no es tan conocida.

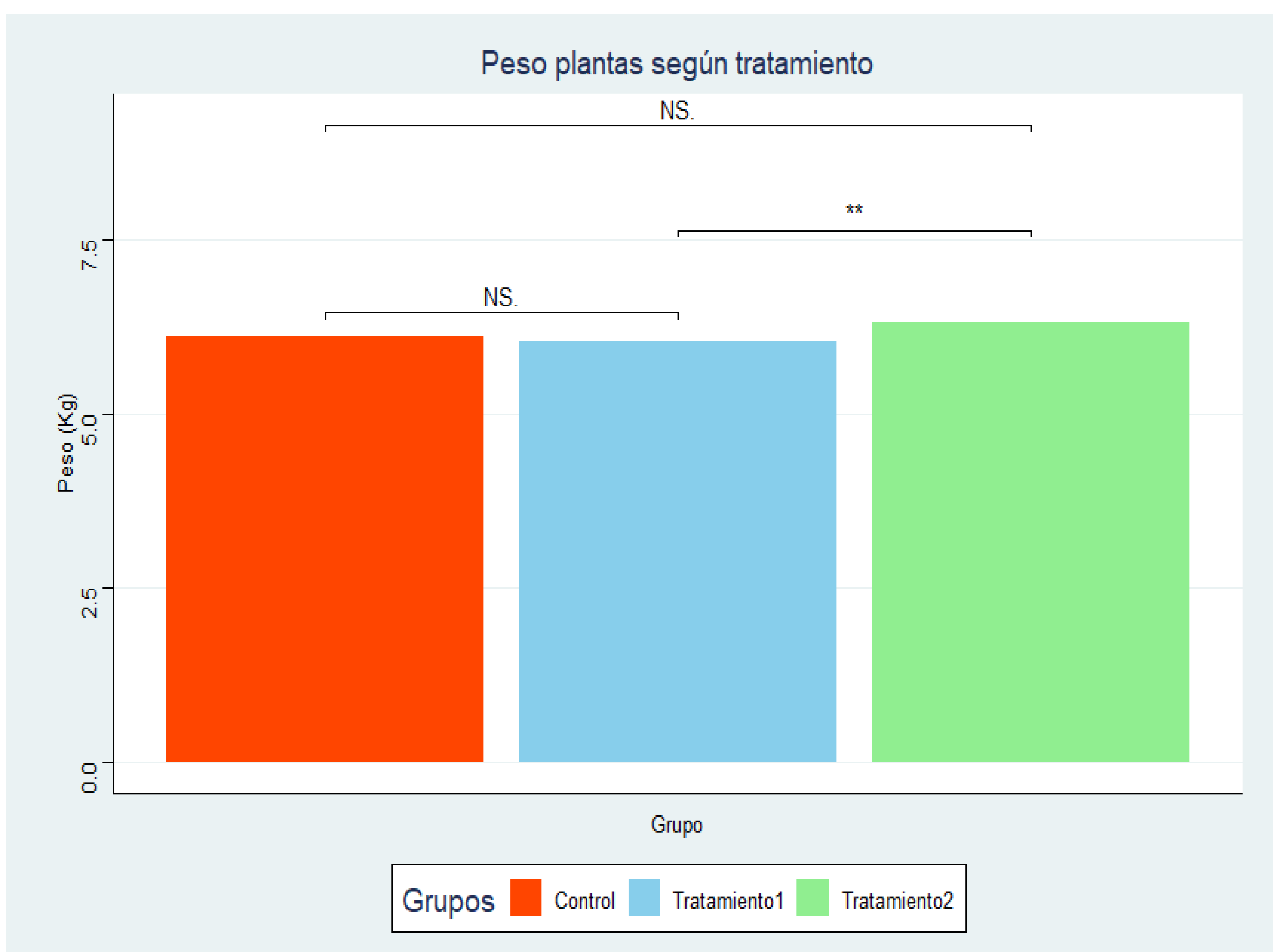


Figura 2. Diagrama de barras mostrando las diferencias en el peso de las plantas según el grupo al que pertenecen. Se puede ver si hay significación (**) o no la hay (NS.) entre los diferentes grupos tras la realización de pruebas post-hoc donde previamente se había realizado un test de ANOVA que sí arrojaba resultados significativos. Gracias a los test post-hoc podemos saber de dónde viene la significación

La base del sistema Bayesiano. Las probabilidades

La estadística Bayesiana es también denominada estadística probabilística. Como su nombre indica se basa en el uso de probabilidades, y no de frecuencias, para obtener las inferencias estadísticas. La probabilidad en este tipo de estadística se define como el grado de creencia o confianza en que se dé un evento. Este grado de creencia puede venir de diversos lados: experimentos previos realizados que hablan de la probabilidad de ese evento, nuestra experiencia propia... es aquí donde esta interpretación de la estadística difiere con la frecuentista puesto que ya no definimos la probabilidad de un evento como el límite de su frecuencia relativa.

Otra diferencia importante entre el sistema probabilístico y el frecuentista reside en que, en el primer sistema, los parámetros del modelo a estudio se consideran desconocidos y sólo podremos obtener de ellos un grado mayor o menor de creencia en su valor. En el segundo sistema, los valores de los parámetros son fijos y nos acercamos a ellos a medida que, siguiendo el ejemplo del lanzamiento del dado mencionado anteriormente, lancemos más veces el dado. No tendría sentido preguntarle a un frecuentista, ¿cuál es la probabilidad de que el dado arroje un 5 en una tirada? Tendría que contestarnos que, aunque no lo sepa, sabe que a medida que repetamos el evento en cuestión, es decir, que arrojemos el dado más veces, podrá saber la frecuencia relativa con la que sale el número 5.

Las dos interpretaciones, la frecuentista y la bayesiana, son válidas y han demostrado numerosas ventajas e inconvenientes.

En otro ejemplo sencillo, si un detective busca saber quién ha cometido un crimen y tiene 4 sospechosos sin ningún tipo de información previa, la posibilidad de cada uno de ellos de ser culpable es del 25 %. En el momento en que puede descartar la culpabilidad de uno tendrá que redistribuir las nuevas credibilidades – de nuevo suponiendo que no hay otra información – de un 33 % al resto de sospechosos.

En la práctica diaria, son tres las diferencias fundamentales que encontramos entre un tipo de estadística y la otra. En la estadística bayesiana tenemos que:

1. Crear el modelo estadístico. Lo cual también ocurre en la estadística frecuentista, pero aquí tenemos que usar **distribuciones** y modelos matemáticos adecuados que nos obligan a tener un conocimiento mínimo sobre ellos. Hoy en día, los programas que se pueden utilizar y que se usan en este póster pueden hacerlo por nosotros bastante bien.
2. Especificar la distribución previa. Lo mismo, requiere de un **conocimiento previo sobre** otros modelos o sobre cómo se comportan **los parámetros en nuestro modelo**. Según nuestra confianza en modelos previos, estas distribuciones serán de una forma u otra. De nuevo, hoy en día, los programas pueden realizar esta labor de forma automática de forma adecuada, eso sí, por defecto, se utilizarán distribuciones “ignorantes” es decir dando por hecho que se sabe poco sobre cómo se comportan los parámetros. El uso de distribuciones previas es un tópico de debate habitual en la estadística bayesiana; ¿debemos usar un modelo ignorante o un modelo una distribución de credibilidades muy determinada? (**figura 3**).

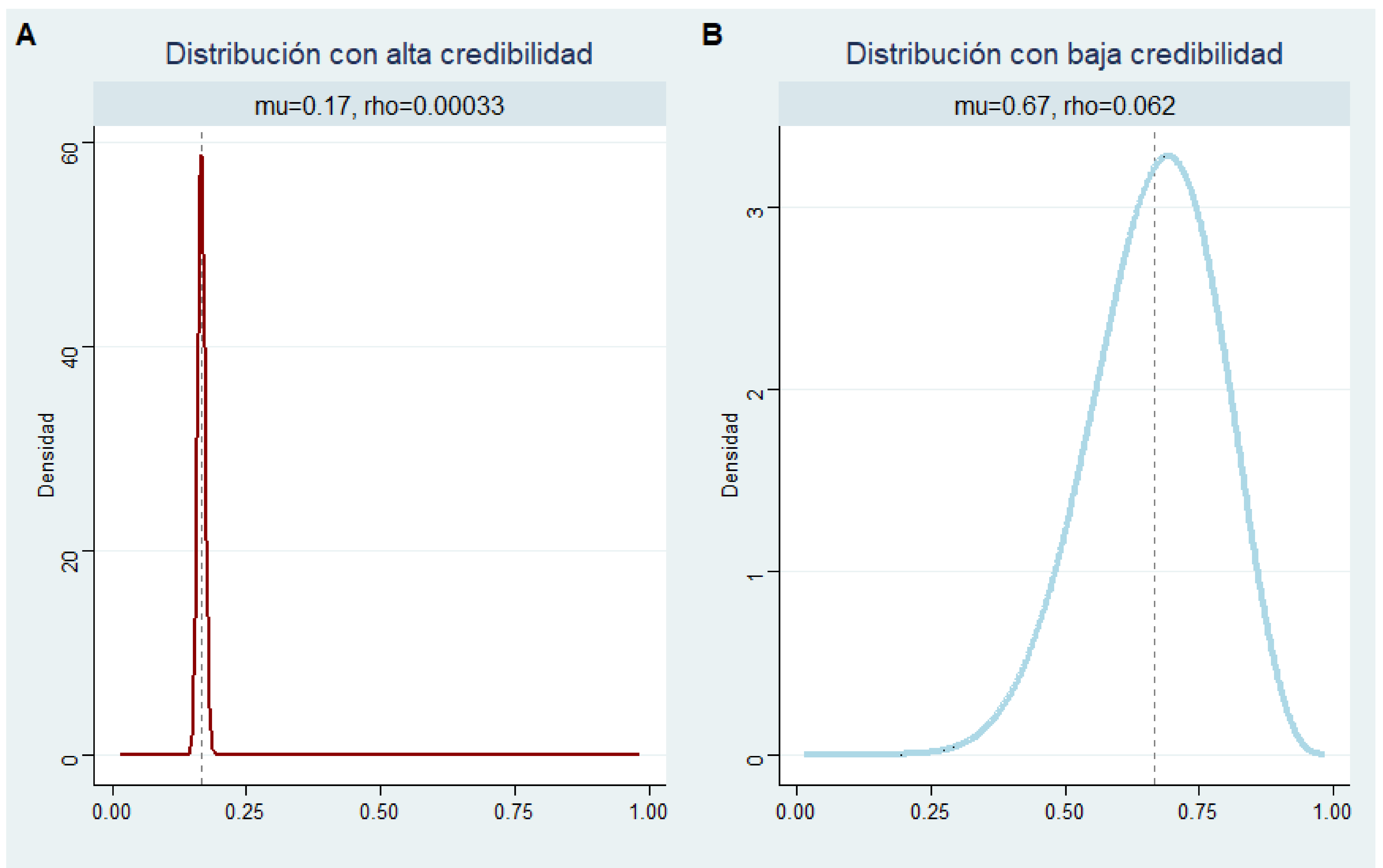


Figura 3. Dos distribuciones: **A** con alta credibilidad (alta certeza en las posibilidades) y **B** con baja credibilidad (baja certeza en las posibilidades).

3. Comprobar que la **distribución posterior** refleja adecuadamente los datos. Un modelo adecuado a los datos es útil, ya que puede predecir resultados con datos futuros o nos permite realizar inferencias sobre nuestra muestra y extrapolarnos a la población general. Si la distribución posterior no es adecuada, tendremos que valorar si ha habido algún error durante el proceso de ejecución del modelo o con el modelo en sí. Este es uno de los pasos más tediosos del proceso puesto que los modelos estadísticos necesitan de tiempo para poderse ejecutar y ver si es adecuado o no (**Figura 4**).

Sistema bayesiano, $\text{beta}(2, 10)$ previa, $s = 15$, $f = 4$

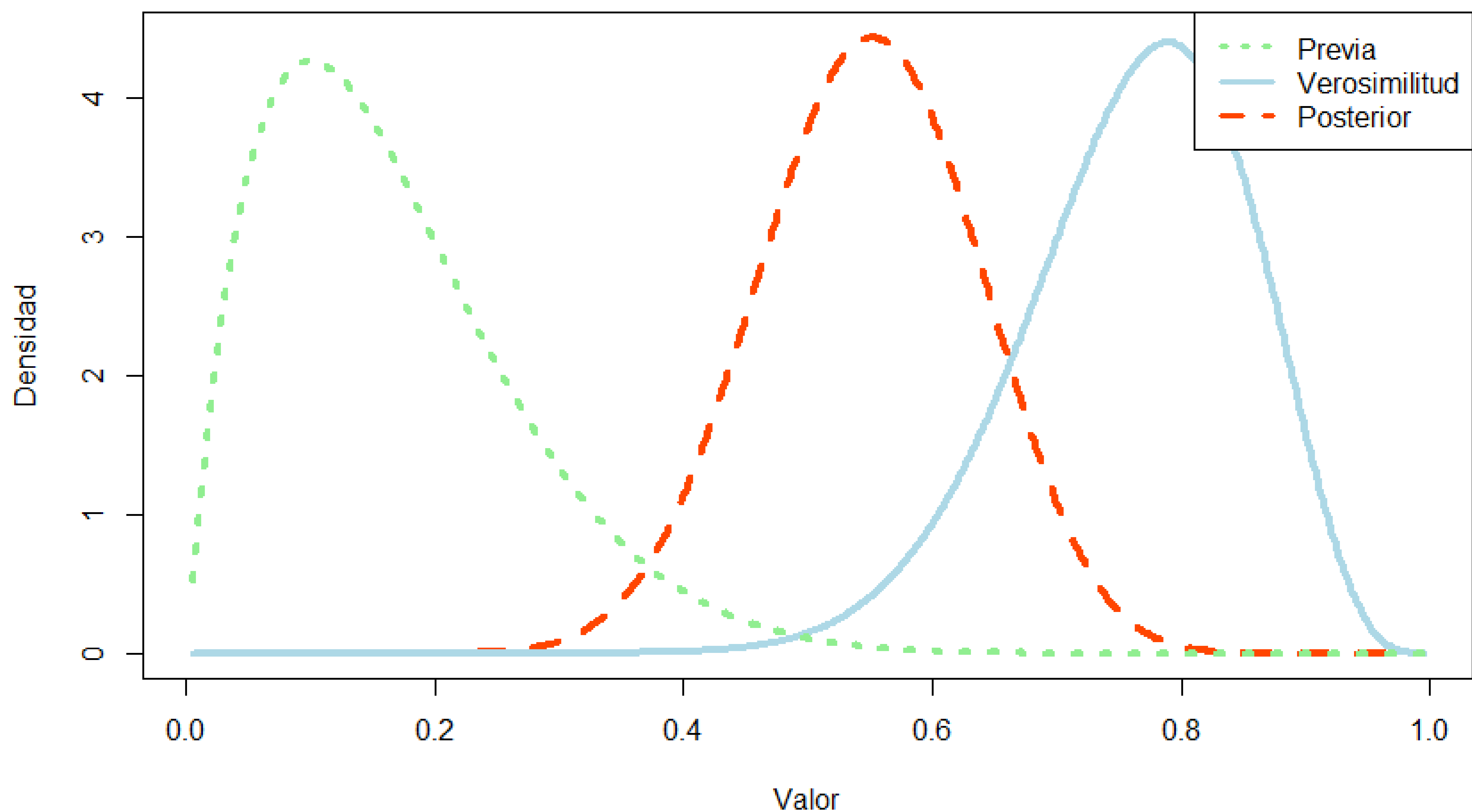


Figura 4. Las tres distribuciones del sistema bayesiano. La distribución previa, la función de verosimilitud y la distribución posterior. En este caso aplicado al ejemplo de tirar una moneda para obtener caras o cruces. La previa – 2 caras de 10; la verosimilitud – 11 caras de 15; la posterior se obtiene de una combinación de las dos anteriores y crea una distribución de credibilidades a las diferentes posibilidades.

Antes de poner en práctica algunos de estos conceptos y empezar a hacer nuestro primer análisis bayesiano de datos vamos a definir algunos conceptos fundamentales de teoría probabilista – algunos de ellos son nuevos y otros se han mencionado de forma más pormenorizada en los párrafos anteriores:

- **Probabilidad condicionada:** Expresa la relación que existe entre dos eventos en los cuales uno de ellos es el evento de interés y el otro un evento que condiciona el previo. Por ejemplo, ¿cuál es la probabilidad de que una lesión focal hepática sea un hepatocarcinoma? Supongamos que es del 1 %. Ahora nos preguntan lo siguiente, ¿y cuál es la posibilidad de que esa lesión sea un hepatocarcinoma teniendo en cuenta que el paciente tiene cirrosis? Aquí es donde entra la probabilidad condicionada y se resuelve teniendo en cuenta la probabilidad de padecer cirrosis teniendo una lesión hepática multiplicado por la probabilidad de tener un hepatocarcinoma dividido entre la probabilidad de tener cirrosis. Expresado en términos sencillos:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Donde $P(A|B)$ es la probabilidad de que suceda A teniendo en cuenta B, $P(B|A)$ es la probabilidad de que suceda B, teniendo en cuenta A, $P(A)$ es la probabilidad marginal de A y $P(B)$ es la probabilidad marginal de B. Este es el llamado **teorema de Bayes**.

El valor del denominador, $P(B)$, no se obtiene mediante un cálculo directo debido a la complejidad computacional que implica. Esto obliga a obtener los valores de las diferentes distribuciones mediante el uso de métodos de la familia **Markov Chain Monte Carlo** (MCMC). Estos comprenden una serie de algoritmos utilizados para obtener una muestra de una distribución de probabilidad. A mayor cantidad de muestras, más cercana será la distribución de la MCMC a la distribución real. Esto ocasiona que cada modelo tenga una mínima variabilidad asociada y que cada vez que se ejecute un modelo con las mismas variables los valores devueltos sean diferentes; no obstante, la diferencia es anecdótica y no tiene un efecto importante.

- **Intervalo de máxima densidad:** Este intervalo, abreviado HDI por sus siglas en inglés (*Highest Density Interval*), expresa que posibilidades de la distribución son más creíbles; es decir aquellos que con el menor número de posibilidades cubren la mayor área bajo la curva. Cuanto más estrecho sea este intervalo, mayor certeza. Para diferenciarlo del intervalo de confianza utilizado en la estadística frecuentista se podría decir que el IC frecuentista expresa la probabilidad de que el intervalo incluya el parámetro verdadero mientras que el HDI expresa que **hay un X % de probabilidades de que el parámetro esté incluido en el intervalo** – se podría discutir que esta última forma, la probabilística, es una forma más natural de entender ese intervalo (**Figura 5**).

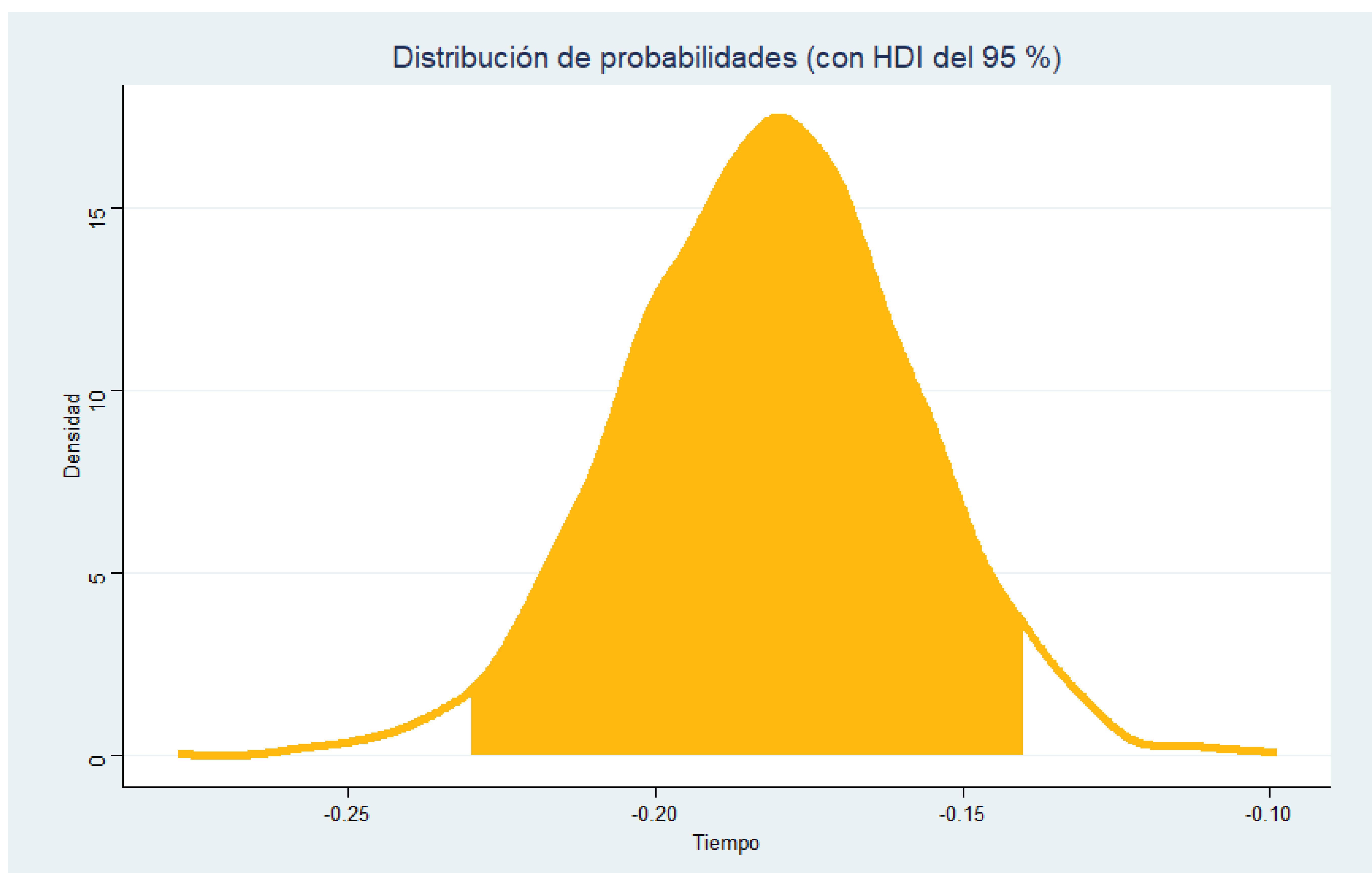


Figura 5. Distribución de densidad del parámetro tiempo (en un estudio de farmacocinética en sangre). Se puede ver como a mayor paso de tiempo menor concentración del fármaco en la sangre. Se ha rellenado el intervalo que contiene con un 95% de probabilidades los valores del parámetro (tiempo) más creíbles. En este caso, el valor único más creíble es el de -0.18.

- **Función de verosimilitud:** Hace referencia a la combinación de las probabilidades de los datos observados. Es una función de los parámetros desconocidos del modelo como una distribución de probabilidades conjuntas. Por ejemplo: si tengo una moneda - que no tiene ningún tipo de alteración y que podemos asumir que las probabilidades de salga cara (1) o cruz (0) son del 50 %, respectivamente – que al tirarla saliera 4 veces seguida cara. La probabilidad de este evento es de 6.25 %. La función de verosimilitud irá cambiando; en este caso se desplazará hacia valores cercanos a 1. Si las siguientes cuatro tiradas son todas cruz la función de verosimilitud cambiará y mostrará una curva más cercana al valor medio real del 50 % de probabilidad de que salga una cara u otra. En la práctica real, no sabemos el valor real del parámetro (en el ejemplo, que la probabilidad de que salga cara o cruz es del 50 %) de ahí la importancia de esta función, teniendo en cuenta los valores obtenidos, que **irá asignado credibilidades a cada una de las posibilidades (Figuras 6 y 7).**

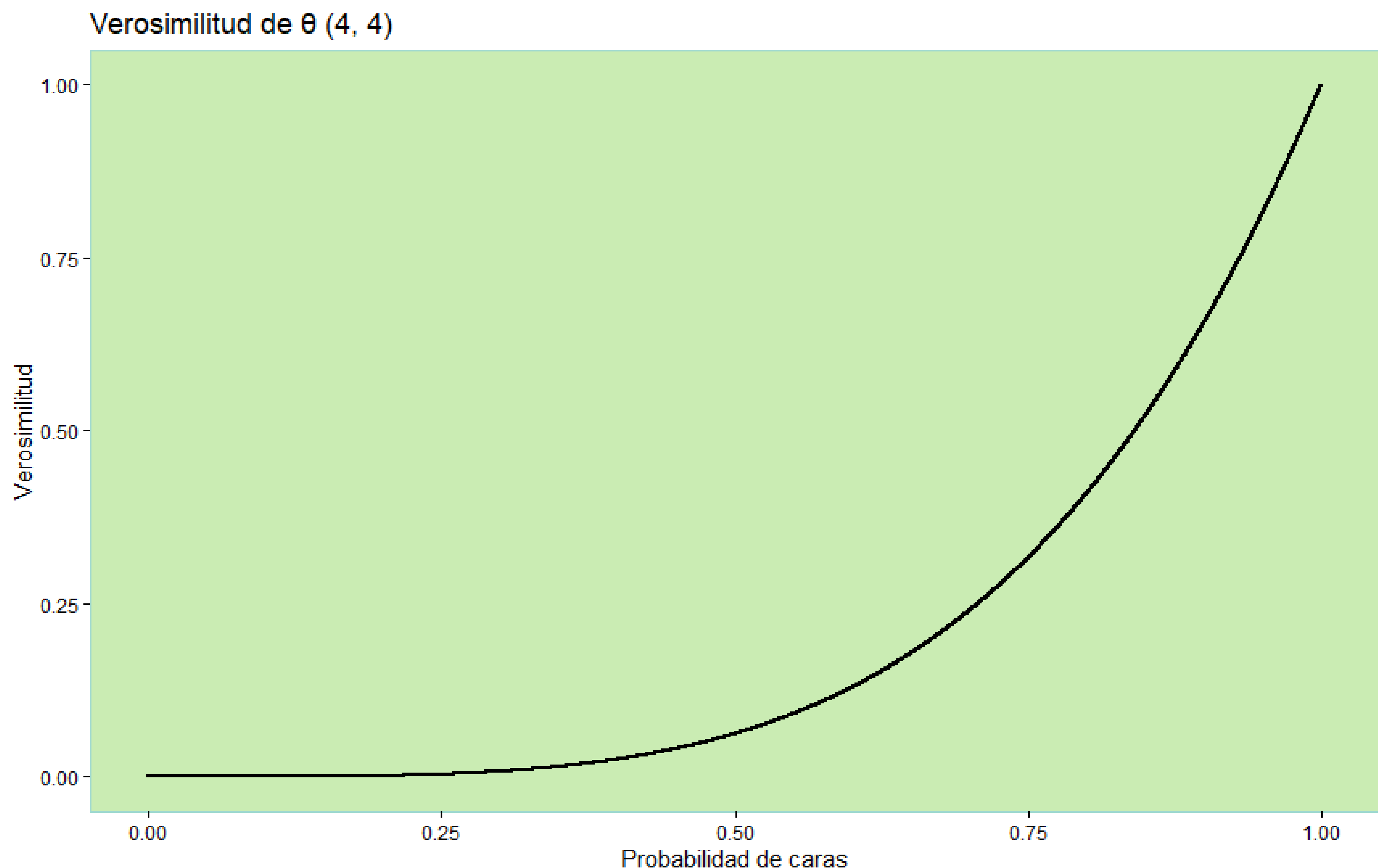


Figura 6. La función de verosimilitud para la probabilidad de cara (1) o cruz (0). Si de 4 tiradas, sacamos cuatro caras. La función muestra que la distribución de probabilidades se acerca a los valores cercanos a 1 en la distribución de probabilidades. Puede que la moneda tenga un sesgo a favor de caras del 80% por ejemplo.

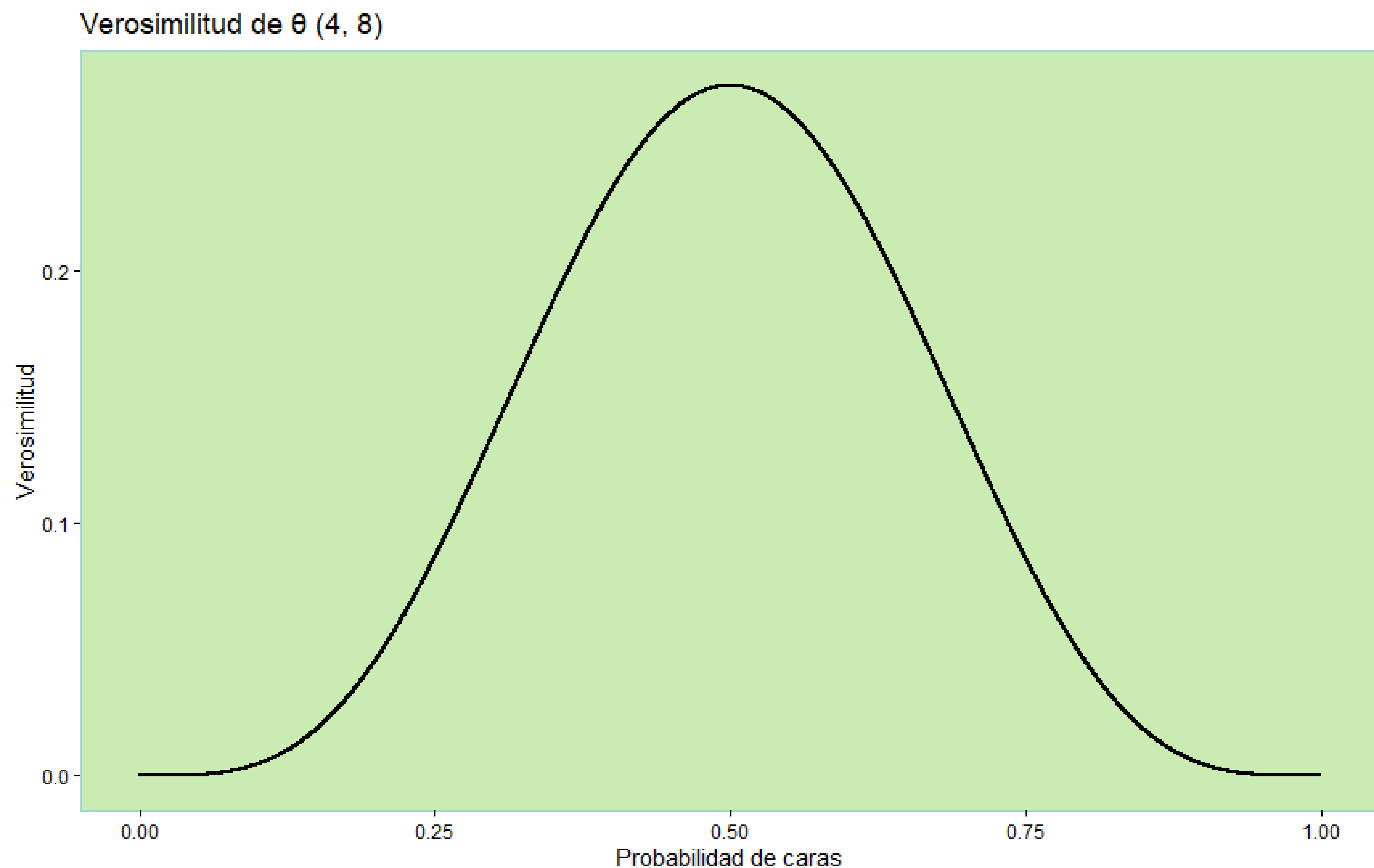


Figura 7. Siguiendo el ejemplo de la gráfica anterior, ahora hemos sacado cuatro cruces. La función de verosimilitud reasigna las credibilidades devolviendo una distribución más centrada. Esta distribución es más acorde con el valor real del parámetro – conocido en este ejemplo de 0.5 – del sesgo de la moneda, representado por la letra θ .

- **Distribución previa:** Distribución de las probabilidades del parámetro a estudio previo al uso de los datos de los que disponemos. ¿Cuál es la probabilidad de que el sol salga mañana por el este? Nuestro conocimiento previo nos diría, con un alto grado de confianza, que la distribución estaría en su totalidad en torno a un intervalo muy estrecho alrededor de 1 (suponiendo en nuestro ejemplo, que 1 equivale a “Sí” y 0 equivale a “No”). Si nuestros datos dicen que esta vez no va a salir por el este sino por el oeste, nuestra evidencia tendría que ser lo suficientemente potente como para desplazar la distribución hacia 0. En este ejemplo, la distribución previa, cuando toda la evidencia indica una credibilidad y nuestros datos otra; indica que serán necesarios más experimentos para corroborar si efectivamente estamos en lo cierto o nuestros datos son una anomalía causada por algún error. A efectos prácticos, y en la mayoría de los casos – ya sea por la dificultad de trasladar los datos de estudios similares al nuestro o por la dificultad de llevar esos datos a nuestro modelo – se utilizan distribuciones previas más o menos *ignorantes*; es decir que no saben demasiado bien donde asignar la certeza.

- **Evidencia / verosimilitud marginal:** Hace referencia a la probabilidad general de los datos obtenidos de acuerdo con el modelo. Es una **función de verosimilitud** que se ha **integrado con todos los valores posibles del parámetro** a estudio – los cuales han sido a su vez **ponderados en base a la credibilidad de dichos valores** para el parámetro.
- **Distribución posterior:** Distribución que se obtiene de la combinación de la distribución previa con la función de verosimilitud dividida entre la evidencia o verosimilitud marginal. Es la distribución que en los análisis Bayesianos va a asignar valores de credibilidad a las diferentes posibilidades de valores del parámetro. Cuanto más “estrecha” sea la curva de distribución mayor certeza tendremos en los valores del parámetro a estudio (**figuras 8 y 9**).

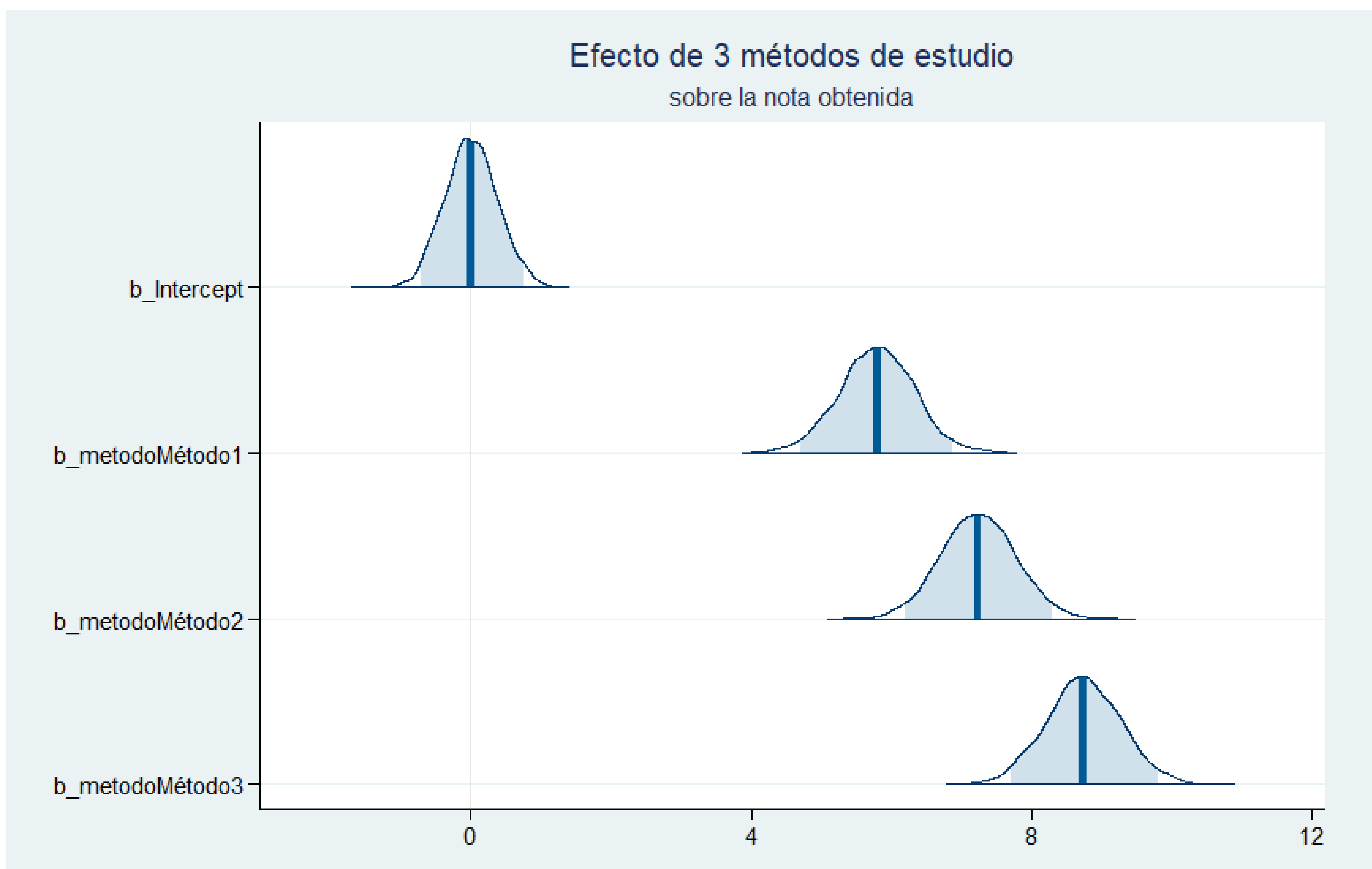
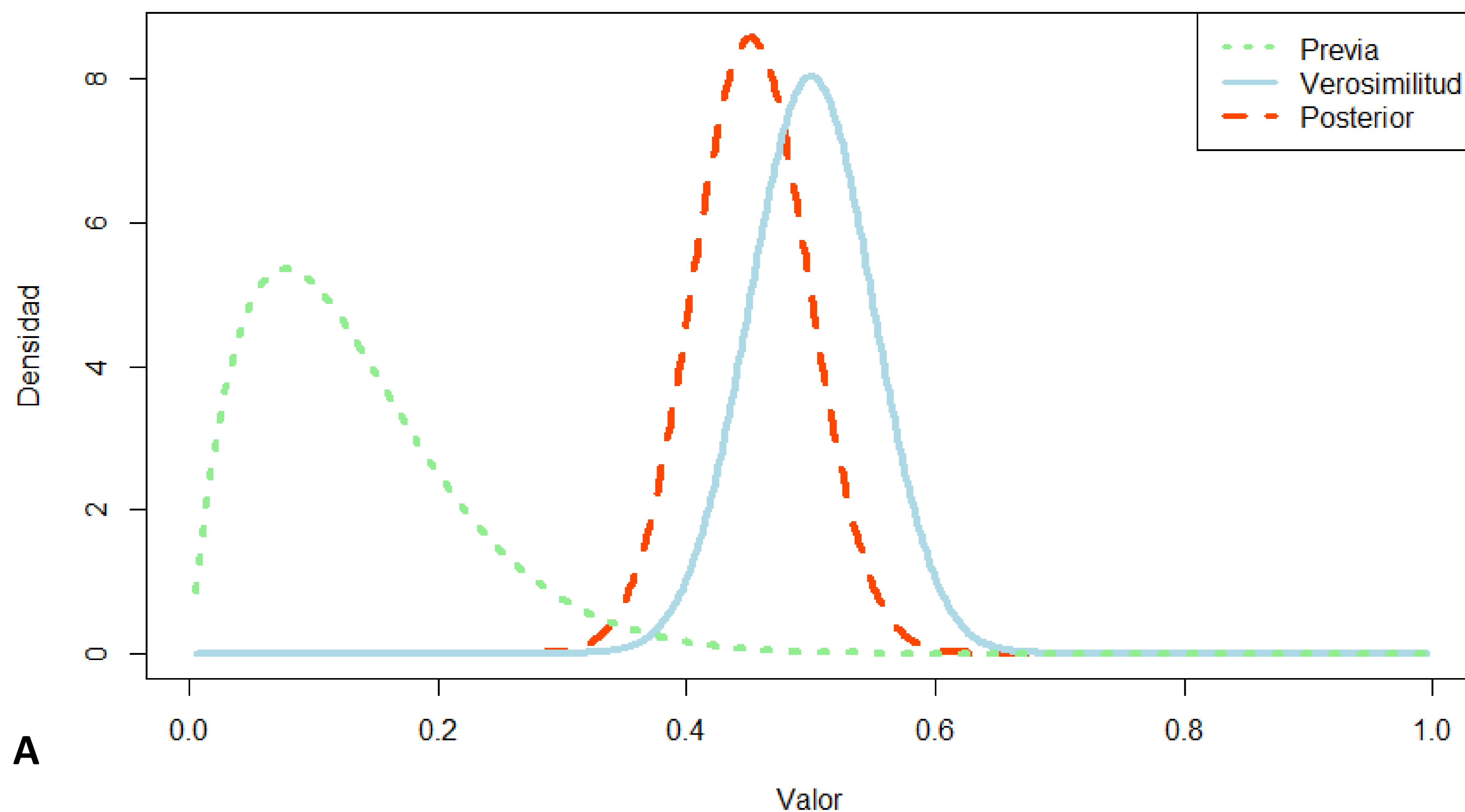


Figura 8. Se representan un total de cuatro distribuciones posteriores con el valor modal representado por la línea azul oscura y el HDI del 95 % por el área en azul claro. Se puede ver como cualquier método, con un grado diferente de certeza supera al método basal (aquí $b_Intercept$) de no estudiar; cada uno obteniendo unas distribuciones diferentes. Si queremos obtener una mejor nota, quizá nos interese el método 3.

Sistema bayesiano con una previa de baja certeza

Previa de 2 / 13

Verosimilitud de 50 / 50

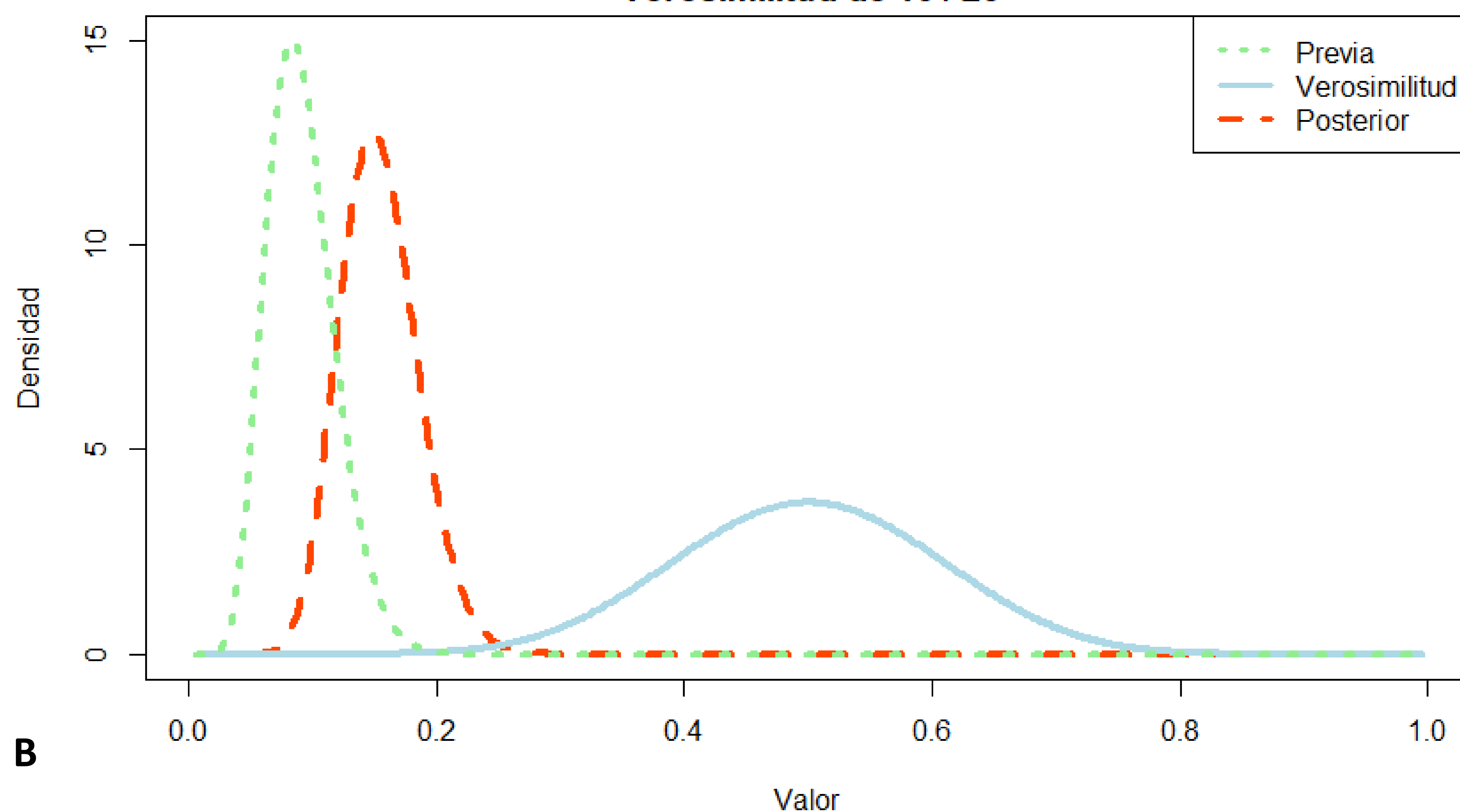


A

Sistema bayesiano con una previa de alta certeza

Previa de 10 / 100

Verosimilitud de 10 / 20



B

Figura 9. Similar en la gráfica de la figura 3. ¿Qué confianza tenemos en que una moneda no está sesgada?. En el caso de la gráfica **A** tenemos poco conocimiento previo de que lo está y mucha evidencia de que la moneda no está sesgada. La distribución posterior vira hacia la verosimilitud, centrándose. En la gráfica **B** ocurre lo contrario, mucha confianza previa en que está sesgada y poca evidencia de que no lo está. La distribución posterior se centraliza muy poco hacia valores de ausencia de sesgo; haría falta más evidencia.

- **Región de equivalencia práctica:** Esta región, abreviada ROPE por sus siglas en inglés (*Region of Practical Equivalence*), hace referencia a un rango dentro de la distribución posterior en la cual se considera que no hay diferencias entre los valores del parámetro obtenidos. Si hablásemos de NHST – sería la hipótesis nula. En un ejemplo donde que no haya diferencias sería el valor 0 se podría aplicar una ROPE de entre -2 y +2 para considerar que esas diferencias de 0 tengan o no un valor importante para nosotros. El rango de este valor, si lo vamos a usar, debería ser escogido con cautela. Para considerar, por tanto, que una distribución posterior es “significativa” – de nuevo intentando aplicar conceptos de la estadística frecuentista a la probabilística – tendríamos que obtener un HDI del 95 %, u otro porcentaje, fuera de la ROPE escogida para poder decir algo parecido a lo que se diría en el caso de NHST de que la hipótesis nula queda rechazada (**figura 10**).

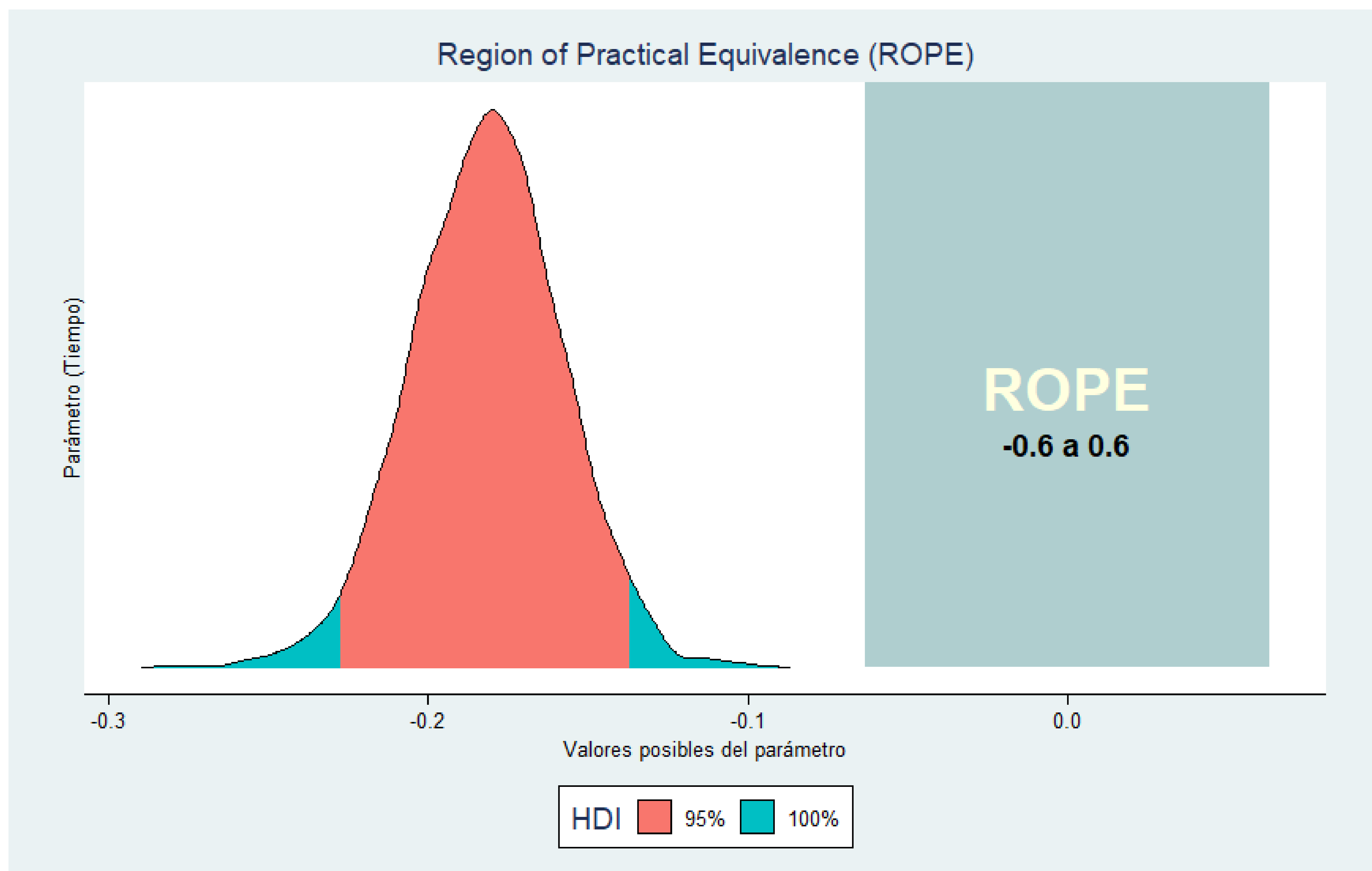


Figura 10. Otra vez el ejemplo de cómo varía un fármaco con el tiempo. Hemos elegido un valor de ROPE de entre -0.6 a 0.6 para considerar ausencia de lavado; en este caso la distribución queda en su totalidad fuera de la ROPE lo cual sería un indicador de que el tiempo influye en la concentración del fármaco.

- **Factor de Bayes** : Llamado BF por sus siglas en inglés (*Bayes Factor*). Es un valor que, a pesar de su nombre, no es muy bayesiano ya que, similar a la ROPE, busca establecer una serie de valores que permitan elegir un modelo sobre otro similar al NHST que compara entre hipótesis alternativa y nula en la estadística frecuentista. **A mayor valor del factor de Bayes, más se favorecerá el primer modelo sobre el segundo y viceversa.** El significado de los valores del Factor de Bayes varía según autores, pero como ejemplo tenemos los de S. Andraszewicz et al (2015) (**Tabla 1**):

Valor del factor de Bayes (M1/M2)	Interpretación
>100	Evidencia extrema para M1
100 – 30	Evidencia muy fuerte para M1
30 – 10	Evidencia fuerte para M1
10 – 3	Evidencia moderada para M1
3 – 1	Evidencia anecdótica para M1
1	No evidencia
1 – 1/3	Evidencia anecdótica para M2
1/3 – 1/10	Evidencia moderada para M2
1/10 – 1/30	Evidencia fuerte para M2
1/30 – 1/100	Evidencia muy fuerte para M2
< 1/100	Evidencia extrema para M2

Tabla 1. Tabla que refleja la interpretación del valor del BF entre dos modelos diferentes para los mismos datos.

Como guía, un modelo será mejor cuanto mayor capacidad para explicar los resultados con la menor cantidad de variables posibles. Una **diferencia clave con el sistema NHST** es que el factor de Bayes nos permite aceptar una hipótesis sobre otra y no únicamente rechazar la hipótesis nula.

Poniendo en práctica estos conocimientos

Cuanto más se sepa cómo funciona algo mejor. De este modo, entenderemos mejor los resultados y podremos detectar si hay algún fallo en las variables, los parámetros o el modelo. No obstante, poder explicar todo el funcionamiento de la estadística bayesiana en tan poco espacio es una tarea imposible, especialmente si se quieren resaltar sus diferencias con la frecuentista. Lo explicado hasta ahora busca servir de *primer* para tratar de entender algunos conceptos de la estadística Bayesiana y tratar de enfocar los problemas estadísticos de una forma diferente.

No obstante, hay muchas cosas de las que no se ha podido hablar como, por ejemplo, ¿cómo se hacen las distribuciones de parámetros?, ¿tiene importancia el tipo de distribución?, ¿qué distribución previa debería elegir? Etc. Se insta al lector, si quiere indagar más sobre el mundo de la estadística Bayesiana que acuda a los libros de referencia y artículos citados en el apartado *Referencias* donde podrá encontrar la respuesta a la mayor parte de sus preguntas y poder ver una explicación más estructurada y de mayor extensión sobre las diferentes peculiaridades de este tipo de estadística.

Vamos a pasar ahora a algo más práctico. ¿Cómo ponemos todo esto en práctica? Afortunadamente, ya no es necesario tantos conocimientos sobre cómo funcionan tantos programas como antes para poder realizar un análisis bayesiano de datos haciendo que este apartado sea más sencillo.



Figura 11. Logos del lenguaje de programación orientado al análisis estadístico R (izquierda) y de la interfaz de desarrollo integrada (IDE) Rstudio. Ambos utilizados en la parte práctica de este trabajo y que se recomiendan a los lectores. Derecha: Logo de R cortesía de Hadley Wickham y otros mediante una licencia CC BY S.A 4.0. Izquierda: Logo de RStudio, dominio público.

En este trabajo, pondremos un ejemplo con el que comenzar, de un análisis estadístico con la perspectiva frecuentista y bayesiana. Se explicarán los pasos a seguir más recomendados con el lenguaje de programación R 4.32 (<https://www.r-project.org/>) y la IDE Rstudio 2023.12.1 Build 402 “Ocean Storm” para Windows ([Download RStudio | The Popular Open-Source IDE from Posit](#)) utilizando varios paquetes añadidos a R destacando el paquete **BRMS** 2.20.4. Además, se utilizarán otros paquetes secundarios para poder analizar mejor los datos visualmente:

- **Brms:** Bürkner P (2017). “brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software*, **80**(1), 1–28. [doi:10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01).
- **Bayesplot:** Gabry J, Mahr T (2024). “bayesplot: Plotting for Bayesian Models.” R package version 1.11.1, <https://mc-stan.org/bayesplot/>.
- **BayestestR:** Makowski D, Ben-Shachar M, Lüdtke D (2019). “bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework.” *Journal of Open Source Software*, **4**(40), 1541. [doi:10.21105/joss.01541](https://doi.org/10.21105/joss.01541), <https://joss.theoj.org/papers/10.21105/joss.01541>.
- **Fishualize:** Nina Schiettekatte, Simon Brandl and Jordan Casey (2019). fishualize: Color Palettes Based on Fish Species. R package version 0.1.0. <https://CRAN.R-project.org/package=fishualize>
- **Ggthemes:** Arnold J (2024). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. R package version 5.1.0.9000, <https://github.com/jrnold/ggthemes>, <http://jrnold.github.io/ggthemes/>, <https://jrnold.github.io/ggthemes/>.
- **LearnBayes:** Albert J P (2018). *Learnbayes: Functions for Bayesian inference*. [CRAN - Package LearnBayes \(r-project.org\)](#)
- **Medicaldata:** Higgins P (2023). *medicaldata: Data Package for Medical Datasets*. <https://higgi13425.github.io/medicaldata/>, <https://github.com/higgi13425/medicaldata/>.
- **Reshape2:** Wickham H. Reshaping Data with the reshape Package. Vol. 21, *Journal of Statistical Software*. 2007. p. 1–20. Available from: <http://www.jstatsoft.org/v21/i12/>
- **Tidyverse:** Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, **4**(43), 1686. [doi:10.21105/joss.01686](https://doi.org/10.21105/joss.01686).

Vamos a empezar entonces. Instalamos R y Rstudio. Una vez instalados en ese orden, abrimos el programa Rstudio y nos aparecerá una imagen como la de la **Figura 12**. No entraremos en demasiados detalles en el uso de este programa en este trabajo. Se insta al lector a remitirse a las diferentes guías, vídeos, libros etc... que hay sobre como trabajar con este programa. Lo primero será apretar **Ctrl + shift + N** para crear un *Script* que nos aparecerá en la parte superior izquierda de la pantalla, desplazando inferiormente a la ventana de *Console*. Todo el código que se va a ir mostrando para este ejemplo se escribe primero en el *Script* y luego se ejecuta dándole al botón de Run que nos aparece en la parte superior derecha de la pantalla *Script* o bien utilizando el atajo **Ctrl + Enter**.

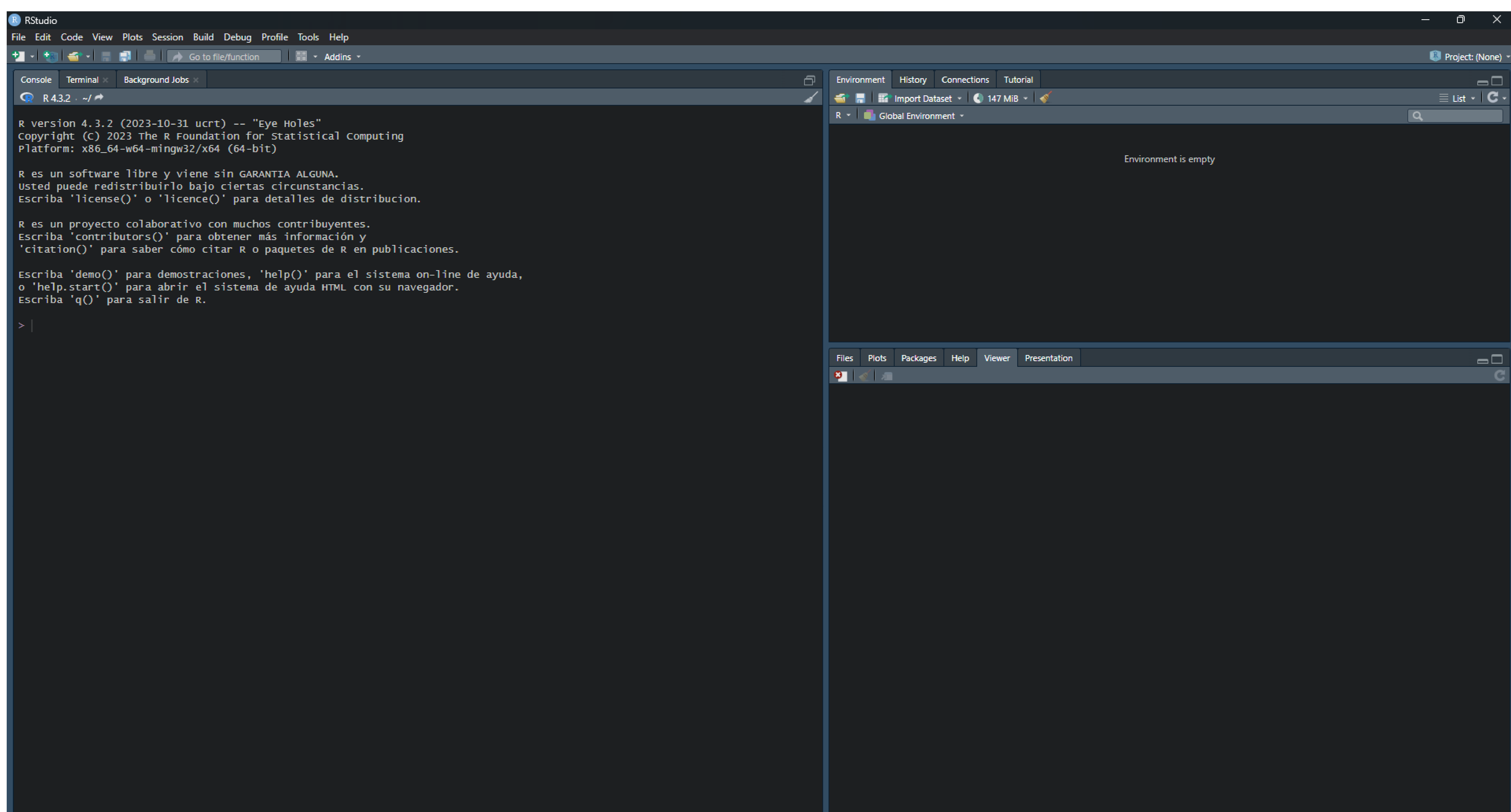


Figura 12. Pantalla que nos muestra RSstudio la primera vez que abrimos el programa. En la pantalla de la izquierda nos aparece la ventana de la consola donde hay unas líneas que hacen referencia al lenguaje de programación R utilizado: versión, licencia etc... Para facilitar el trabajo utilizaremos Scripts que se pueden abrir utilizando el atajo de teclado Ctrl + shift + N o clicando en el botón de arriba a la izquierda que es una hoja blanca con un “+” verde por encima y dándole a la opción *R Script* en las opciones que nos muestra. La ventana superior derecha nos muestra el *Environment*, es decir, todo lo que vamos creando en el *Script*: variables, bases de datos, objetos u otros objetos. En la ventana inferior derecha tenemos varias pestañas útiles la de *Plots* nos muestra el gráfico que hayamos ejecutado – y anteriores –, la opción de *Help* por si necesitamos ayuda y el de *Packages* que nos muestra todos los paquetes que tengamos instalados.

Primero de todo hay que instalar los paquetes mencionados. Esto solo hace falta realizarlo la primera vez, luego solo habrá que llamarlos cada vez que abrimos el *Script* para utilizar sus funciones. Llamarlos solo requiere de unos pocos segundos mientras que la instalación puede alargarse algunos minutos dependiendo del tamaño de la instalación:

```
packages <- c("bayesplot", "bayestestR", "brms", "coda", "cowplot",  
"cubelyr", "devtools", "fishualize", "GGally", "ggdist", "ggExtra",  
"ggforce", "ggmcmc", "ggthemes", "janitor", "lisa", "loo",  
"medicaldata", "palettetown", "patchwork", "psych", "remotes", "reshape2",  
"rstan", "santoku", "scales", "scico", "standarize", "tidybayes",  
"tidyverse")  
  
install.packages(packages, dependencies = T)
```

Además, también instalaremos los siguientes paquetes desde repositorios de *github* puesto que aún no están disponibles en los servidores oficiales de R (CRAN):

```
remotes::install_github("clauswilke/colorblindr")  
devtools::install_github("dill/beyonce")  
devtools::install_github("ropenscilabs/ochRe")
```

Una vez instalados los paquetes, código que se puede poner tanto en el *Script* como en la consola puesto que solo utilizaremos una vez – hay que recordar pues de quitarlo del *Script* para no ejecutarlo accidentalmente. Dejaremos el siguiente código al inicio del *Script* donde llamamos a los paquetes que hemos instalado y vamos a utilizar:

```
library(brms) # paquete que permite realizar el análisis bayesiano  
library(tidybayes)  
library(medicaldata) # de donde obtendremos la base de datos  
library(bayestestR)  
library(bayesplot)  
library(fishualize) # para elegir colores  
library(reshape2)  
library(tidyverse)
```

Hay que tener en cuenta que todo lo que queda a la derecha del símbolo # en la misma línea no se lee en la consola; es utilizado habitualmente cuando se quiere aclarar que es lo que hace el código al otro lado de la #. Vamos a utilizar una base de datos contenida en el paquete *medicaldata*. Para ello utilizamos el siguiente código:

```
df <- polyps #la variable df contiene datos ahora  
glimpse(df) #Nos permite echar un vistazo a esta base de datos
```

Lo cual nos devolverá el siguiente resultado:

```
Rows: 22  
Columns: 7  
$ participant_id <chr> "001", "002", "003", "004", "005", "006", "007",  
"008", "009", "010", "011", "012", "013", "014"..  
$ sex <fct> female, female, male, female, male, female, female,  
male, male, male, male, female, male, male, ...  
$ age <dbl> 17, 20, 16, 18, 22, 13, 23, 34, 50, 19, 17, 23, 22,  
30, 27, 23, 22, 13, 34, 23, 22, 42  
$ baseline <int> 7, 77, 7, 5, 23, 35, 11, 12, 7, 318, 160, 8, 20, 11,  
24, 34, 54, 16, 30, 10, 20, 12  
$ treatment <fct> sulindac, placebo, sulindac, placebo, sulindac,  
placebo, sulindac, placebo, placebo, placebo, su..  
$ number3m <int> 6, 67, 4, 5, 16, 31, 6, 20, 7, 347, 142, 1, 16, 20,  
26, 27, 45, 10, 30, 6, 5, 8  
$ number12m <dbl> NA, 63, 2, 28, 17, 61, 1, 7, 15, 44, 25, 3, 28, 10,  
40, 33, 46, NA, 50, 3, 1, 4
```

Nos muestra la consola que tenemos un total de 22 filas (casos) y 7 variables. En este caso se trata de una base de datos donde se quiere analizar el efecto del uso del fármaco *sulindac* para reducir el número de pólipos de colon en pacientes con poliposis adenomatosa familiar. Las **variables** son el número de identificación, el sexo, la edad, el número de pólipos en el momento basal, a los 3 meses, a los 12 meses y el tratamiento que han recibido.

Ahora vamos a dejar los datos más sencillos para nosotros. Vamos a traducir los nombres de las variables para que nos sea más fácil leerlas con el siguiente código:

```
names(df) <- c("ID", "Sexo", "Edad", "Basal", "Tratamiento", "tresMeses",  
"doceMeses") #Cambia el nombre de variables  
df$Sexo <- factor(df$Sexo, levels = c("female", "male"), labels = c("Mujer",  
"Hombre")) #Cambia los niveles de la variable Sexo  
df$Tratamiento <- factor(df$Tratamiento, levels = c("placebo", "sulindac"),  
labels = c("Placebo", "Sulindac")) #Cambia los niveles de la variable  
Tratamiento  
head(df) #Otra función para ver el df, esta vez solo los 6 primeros valores
```

Ahora el código nos arrojará:

	ID	Sexo	Edad	Basal	Tratamiento	tresMeses	doceMeses
1	001	Mujer	17	7	Sulindac	6	NA
2	002	Mujer	20	77	Placebo	67	63
3	003	Hombre	16	7	Sulindac	4	2
4	004	Mujer	18	5	Placebo	5	28
5	005	Hombre	22	23	Sulindac	16	17
6	006	Mujer	13	35	Placebo	31	61

Si quisiéramos realizar un análisis frecuentista. Digamos que queremos ver el efecto sobre los pólipos a los doce meses de las variables Sexo, Edad, Basal y Tratamiento utilizaríamos el siguiente código:

```
modelo_fr <- lm(doceMeses ~ Sexo + Edad + tresMeses + Basal + Tratamiento,  
data = df)  
summary(modelo_fr) #Para ver los resultados del modelo
```

Este código nos permite obtener un modelo donde podremos analizar como influyen las variables independientes sobre la dependiente – en este caso el número de pólipos a los doce meses. La segunda línea de código nos muestra el siguiente resultado:

Call:

```
lm(formula = doceMeses ~ Sexo + Edad + Basal + tresMeses + Tratamiento,  
    data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.904	-8.129	-1.880	7.800	22.270

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	43.7953	11.8434	3.698	0.002387	**
Sexomale	2.5633	7.3141	0.350	0.731211	
Edad	-0.5347	0.4221	-1.267	0.225905	
Basal	1.1515	0.4555	2.528	0.024122	*
tresMeses	-1.0359	0.4297	-2.411	0.030231	*
TratamientoSulindac	-34.5149	7.0158	-4.920	0.000226	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.25 on 14 degrees of freedom

(2 observations deleted due to missingness)

Multiple R-squared: 0.7025, Adjusted R-squared: 0.5963

F-statistic: 6.612 on 5 and 14 DF, p-value: 0.002338

No vamos a entrar en análisis de interacciones entre variables para mantener el ejemplo sencillo, pero tampoco tienen mucho peso en este caso.

Los resultados nos muestran varias cosas:

- Las variables influyen de forma significativa en el modelo globalmente (F-Statistic) y podemos ver que esto se debe a las variables:
 - Basal: En sentido positivo siendo una variable continua, es decir, si teníamos más pólipos basalmente, más tendremos a los 12 meses.
 - tresMeses: En sentido negativo siendo una variable continua, si teníamos a los 3 meses menos pólipos, menos tendremos a los 12 meses.
 - TratamientoSulindac: En sentido negativo como variable categórica, es decir, si el paciente tomaba Sulindac, tendrá a los 12 meses menos pólipos, este además parece ser la variable que más importancia tiene con un *Estimate* de -34,5149.

- El error residual es de 13.25, dentro de lo aceptable para la magnitud de la variable dependiente.
- Los valores de la R^2 y de la R^2 ajustada son aceptables.

Si queremos visualizar los resultados y ver la evolución de los pólipos según el tratamiento a lo largo del tiempo utilizaremos:

```
df2 <- melt(df, id.vars = c("ID", "Sexo", "Edad", "Tratamiento"),  
variable.name = c("Tiempo"), value.name = "Polipos") #Un data.frame nuevo  
para que lea los datos correctamente la función ggplot de creación de  
gráficos  
plot_2 <- ggplot(data = df2, mapping = aes(y = Polipos, fill = Tratamiento))  
plot_2 + geom_boxplot() + facet_wrap(~Tiempo, labeller = labeller(names)) +  
ylim(c(0,80)) + ggthemes::theme_stata(base_size = 12) + guides(fill =  
guide_legend(title = "Tratamiento")) + ylab("Pólipos")
```

Nos mostrará:

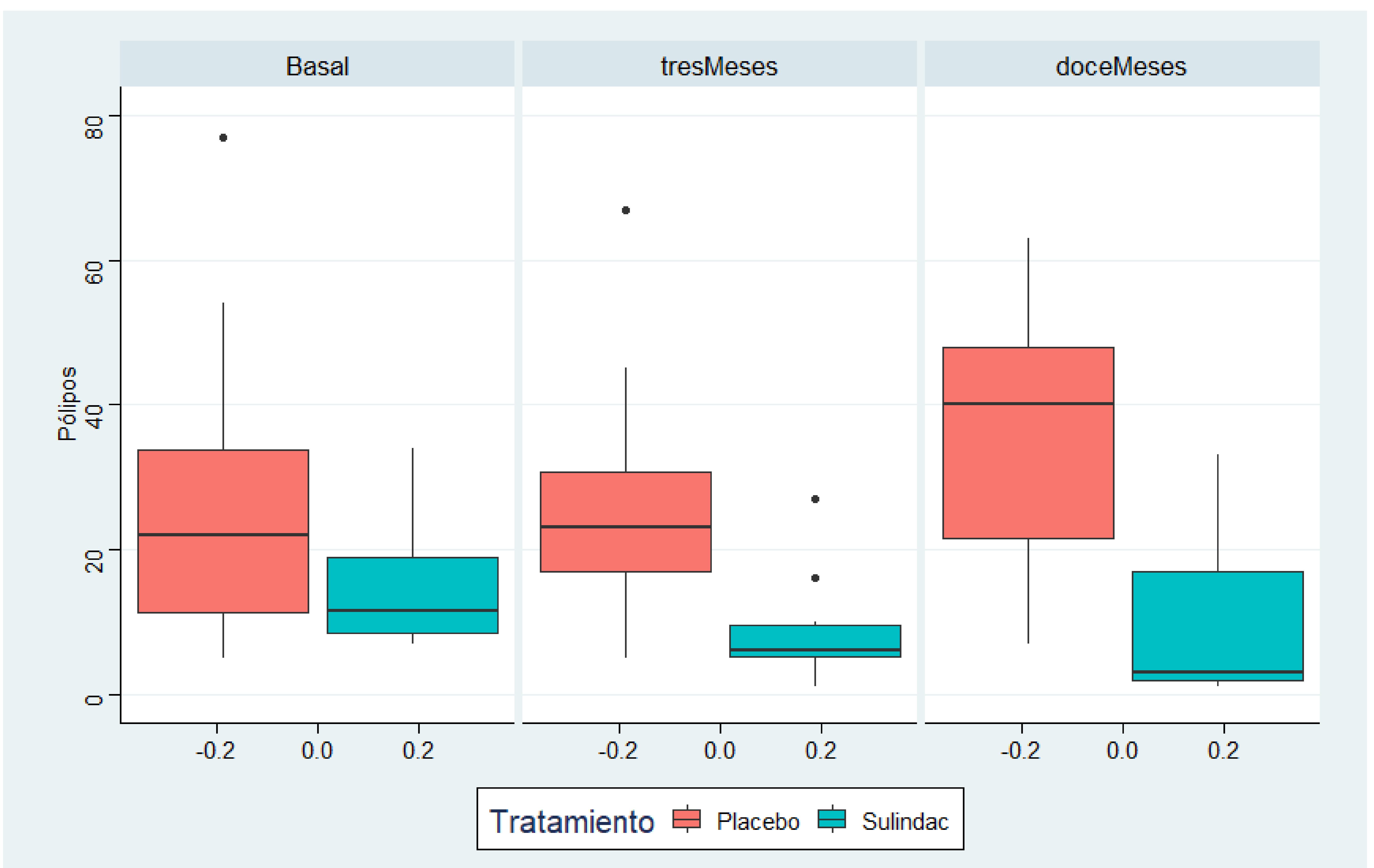


Figura 13. Diagrama de cajas que muestra la evolución de los pólipos en el momento basal, a los tres meses y a los doce meses según el tratamiento.

Ahora vamos a probar un modelo Bayesiano para explicar y tratar de analizar los resultados obtenidos. En este caso el código que tenemos que utilizar es el siguiente:

```
fit <- brm(doceMeses ~ Sexo + Edad + Basal + tresMeses + Tratamiento,  
family = gaussian(), data = df, save_all_pars = TRUE) #Para realizar el modelo  
Bayesianofit  
fit
```

Se podrá comprobar la diferencia de tiempo entre realizar un análisis Bayesiano relativamente sencillo – nos tarda de varios segundos a un minuto - y ejecutar una regresión lineal frecuentista – que tarda menos de 1 segundo. La segunda línea de código mostrará los resultados:

```
Family: gaussian  
Links: mu = identity; sigma = identity  
Formula: doceMeses ~ Sexo + Edad + Basal + tresMeses + Tratamiento  
Data: df (Number of observations: 20)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
total post-warmup draws = 4000  
  
Population-Level Effects:  
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
Intercept          44.09    12.78    18.25    69.46 1.00    3306    2971  
SexomHombre         2.47     8.20   -14.18    18.73 1.00    2718    2524  
Edad                -0.54     0.46    -1.45     0.36 1.00    2828    2741  
Basal               1.14     0.50     0.17     2.13 1.00    1465    2077  
tresMeses          -1.03     0.47    -1.98    -0.12 1.00    1474    2008  
TratamientoSulindac -34.54     7.77   -50.03   -19.01 1.00    2200    2545  
  
Family Specific Parameters:  
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
sigma         14.33     3.00     9.88    21.49 1.00    1995    1987
```

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

Se puede ver como los valores son similares a los obtenidos en el modelo frecuentista. Pero aquí no encontramos que se nos hable de p valor ni de R^2 . Aunque si nos aparece el valor de CI (*Confidence Interval*) que, recordemos, no es lo mismo que el HDI que utilizaremos:

```
hdi(fit, ci = c(0.89, 0.95)) #Nos devuelve el HDI
```

Lo cual nos devuelve:

Parameter	89% HDI	95% HDI
(Intercept)	[22.98, 63.34]	[17.86, 68.54]
SexoHombre	[-10.12, 15.60]	[-14.03, 18.88]
Edad	[-1.28, 0.17]	[-1.41, 0.41]
Basal	[0.37, 1.92]	[0.20, 2.15]
tresMeses	[-1.79, -0.32]	[-1.98, -0.13]
TratamientoSulindac	[-46.20, -21.23]	[-49.79, -18.91]

Se puede comprobar los valores del rango en los que probablemente esté el valor real de cada uno de los parámetros. Se puede ver como el sexo masculino pasa por el valor 0, por tanto, probablemente no haya diferencia entre sexo masculino y femenino. Lo mismo ocurre para edad. No obstante, para las variables *Basal*, *tresMeses* y *TratamientoSulindac* esto no ocurre, indicando que probablemente sí que tengan una influencia sobre la variable dependiente.

Si queremos ver esto de forma gráfica podemos utilizar el siguiente código:

```
result <- hdi(fit, ci = c(0.95, 0.89))plot(result) +  
ggthemes::theme_stata() + xlab("Posibles valores del parámetro") +  
yaxis_text(on = TRUE, angle = 0) + ylab("Parámetros")
```

Que nos devuelve la **figura 13** donde se ve gráficamente la distribución posterior donde están coloreados de azul, amarillo y rojo los intervalos de HDI de 100 %, 95 % y 89 % respectivamente. Se puede ver como la variable *TratamientoSulindac* queda totalmente alejada de 0, mientras que el 95 % de los valores creíbles de los parámetros *Basal* y *tresMeses* también quedan fuera.

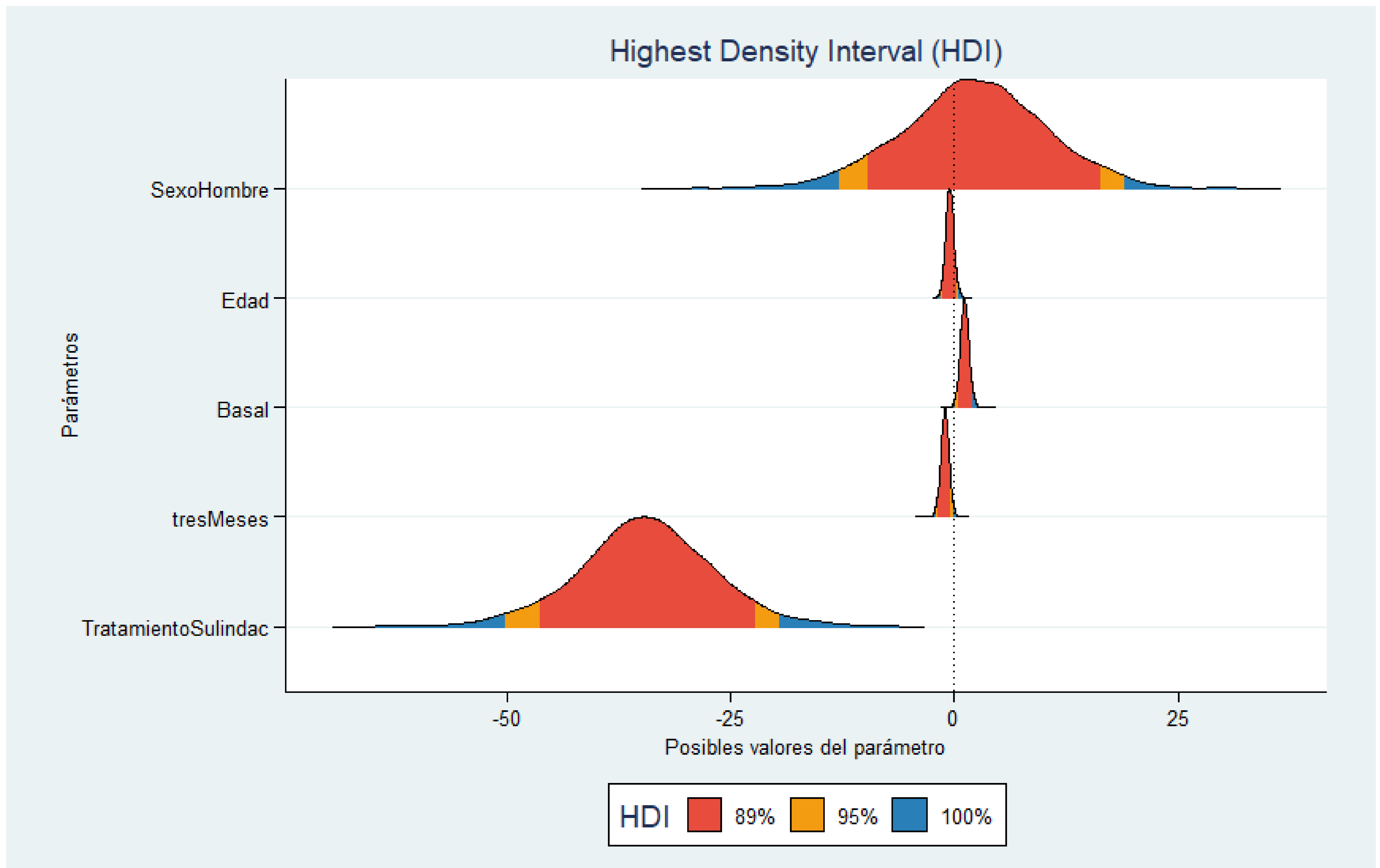


Figura 14. Distribuciones posteriores de los diferentes parámetros a estudio para ver su influencia sobre la variable que representa el número de pólipos colónicos a los 12 meses. Se puede ver como el último parámetro es el que más influencia tiene mientras que los valores *Basal* y *tresMeses* también excluyen el 95% de sus valores creíbles del valor 0.

Podemos inferir que son tres las variables que influyen sobre el número de pólipos a los 12 meses. Las variables *Edad* y *Sexo* no parecen tener mucha influencia. Por el contrario, las variables *Basal*, *tresMeses* y *TratamientoSulindac* sí influyen. No obstante, ¿es el valor de *Basal* y *tresMeses* realmente significativo? ¿Qué tengamos 1 pólipo de más o de menos realmente tiene un impacto clínico?, o ¿un impacto clínico relevante? Esto puede ser motivo de discusión y debate. Utilizando la ROPE (región de equivalencia práctica) podemos tratar de orientarnos. El valor de esta ROPE es algo subjetivo y puede variar con el investigador. Para este ejemplo vamos a aceptar el rango que nos ofrezca la herramienta estadística como óptimo por mantener la simplicidad. Introducimos el siguiente código:

```
rope(fit, ci_method = "HDI", range = "default")
```

Que nos devolverá los siguientes resultados:

```
# Proportion of samples inside the ROPE [-2.08, 2.08]:
```

Parameter	inside ROPE
Intercept	0.00 %
SexoHombre	20.99 %
Edad	100.00 %
Basal	99.13 %
tresMeses	100.00 %
TratamientoSulindac	0.00 %

El valor de ROPE que la herramienta nos recomienda está en la primera línea: entre -2.08 y 2.08 haciendo que las variables *Edad*, *Basal* y *tresMeses* tengan su totalidad o práctica totalidad dentro de esta área. *Sexomale* tiene un 21 % dentro del ROPE que también se puede considerar demasiado. La única variable que no está totalmente incluida es la de *TratamientoSulindac*. Estos resultados orientan a que sea esta última variable la que realmente tenga más influencia y, sobre todo una influencia importante. Puede que las variables *Basal* y *tresMeses* sean “significativas” pero su impacto real es muy limitado.

Con el siguiente código podremos representar la ROPE gráficamente con todas las variables (**Figura 14**):

```
plot(rope(fit, ci_method = "HDI", range = "default")) + theme_replace() +  
xlab("Valores posibles de los parámetros") + ylab("Parámetros") +  
ggtitle("Región de Equivalencia Práctica (ROPE)") + guides(fill =  
guide_legend("HDI"))
```

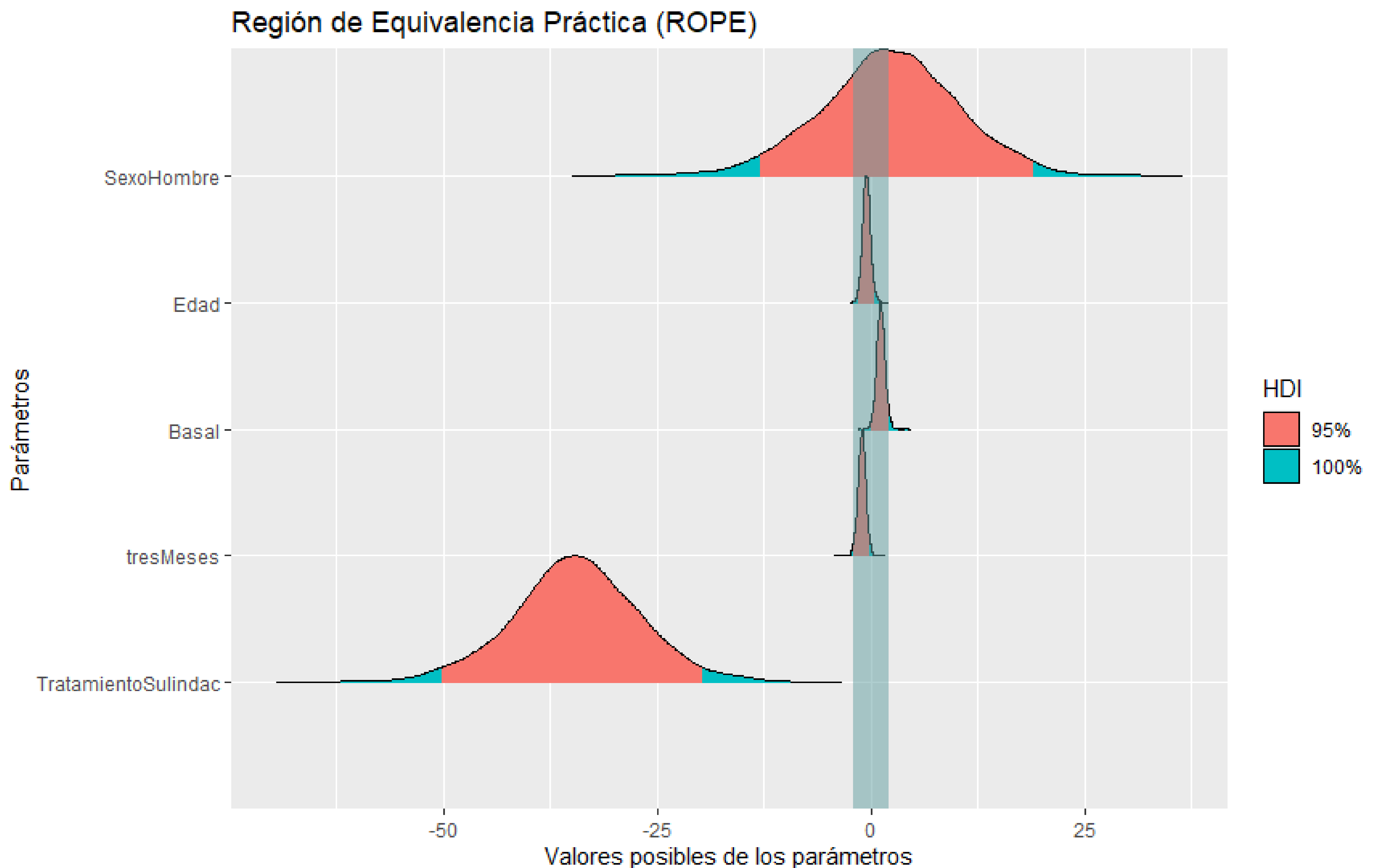


Figura 15. Distribuciones posteriores de los diferentes parámetros a estudio con la ROPE – región de equivalencia práctica reflejada en azul claro alrededor de 0. Entre los valores -2.08 y $+2.08$ consideramos que hay una equivalencia y que, básicamente, pequeñas variaciones, aunque reales – utilizando el jergón de la estadística frecuentista – no tienen una significación.

Con estos resultados vamos perfilando unas inferencias. Aunque tengan una influencia real las variables – *Basal*, *tresMeses* y *TratamientoSulindac*; únicamente parece tener una influencia real esta última. Ahora, por último, trataremos el Factor de Bayes, el cual nos permitirá comparar dos (o más) modelos diferentes para ver cual se puede considerar mejor o más completo.

Para ello, vamos a realizar otro modelo, en este caso, por simplicidad vamos a eliminar un valor como por ejemplo la edad, que parece no tener mucha influencia sobre el valor final y por tanto se puede considerar una variable redundante:

```
fitsinEdad <- brm(doceMeses ~ Sexo + Basal + tresMeses + Tratamiento,  
family = gaussian(), data = df, save_all_pars = TRUE)  
fitsinEdad
```

Al ejecutar la segunda línea nos devuelve:

```
Family: gaussian
Links: mu = identity; sigma = identity
Formula: doceMeses ~ Sexo + Basal + tresMeses + Tratamiento
Data: df (Number of observations: 20)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000

Population-Level Effects:
              Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept          30.74     6.71   17.21   43.72 1.00   3065   2664
SexoHombre          -1.50     7.38  -15.75   13.24 1.00   2496   2366
Basal                1.30     0.49    0.35    2.27 1.00   1398   1984
tresMeses           -1.15     0.47   -2.07   -0.25 1.00   1377   1797
TratamientoSulindac -33.65     7.87  -49.24  -18.67 1.00   2042   2383

Family Specific Parameters:
              Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma          14.42     2.76   10.21   20.88 1.00   2191   2462

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Donde podemos ver algunos pequeños cambios en los valores de cada una de las variables, aunque globalmente las conclusiones son prácticamente las mismas. Para poder comparar ambos modelos es fundamental que en cada uno de los modelos que hagamos aparezca escrito: `save_all_pars(TRUE)`. Escribimos entonces:

```
bayes_factor(fit, fitsinEdad)
```

Que nos devolverá lo siguiente:

```
Iteration: 1  
Iteration: 2  
Iteration: 3  
Iteration: 4  
Iteration: 5  
Iteration: 6  
Iteration: 1  
Iteration: 2  
Iteration: 3  
Iteration: 4  
Iteration: 5  
Estimated Bayes factor in favor of fit over fitsinEdad: 2.44039
```

Por tanto, siguiendo los valores de Andraszewicz et al. (2015) diríamos que el modelo con la variable *Edad* tienen una mayor evidencia anecdótica que el modelo sin la variable *Edad*.

Si hacemos otro modelo, en este caso sin la variable *tresMeses*:

```
fitsinTresMeses <- brm(doceMeses ~ Sexo + Edad + Basal + Tratamiento,  
family = gaussian(), data = df, save_all_pars = TRUE)
```

Y ejecutamos el siguiente código:

```
bayes_factor(fit, fitsinTresMeses)
```

Nos devuelve:

```
Iteration: 1  
Iteration: 2  
Iteration: 3  
Iteration: 4  
Iteration: 5  
Iteration: 6  
Iteration: 1  
Iteration: 2  
Iteration: 3  
Iteration: 4  
Iteration: 5  
Estimated Bayes factor in favor of fit over fitsinTresMeses: 13.70528
```

Que implicaría una evidencia a favor del modelo con la variable 3 meses sobre el que no la tiene que la podemos calificar de fuerte. Es decir, esta variable es importante y útil para el modelo, aunque pueda no tener una gran relevancia clínica.

El valor exacto del factor de Bayes obtenido mediante esta función varía ligeramente, aunque es estable. Esto es debido a que se infiere cada vez de una muestra similar pero diferente.

37 Congreso
Nacional
CENTRO DE
CONVENCIONES
INTERNACIONALES

Barcelona
22/25
MAYO 2024

seram
Sociedad Española de Radiología Médica

FERM
FUNDACIÓN ESPAÑOLA DE RADIOLOGÍA MÉDICA

RC | RADIOLEGS
DE CATALUNYA

Conclusión

Conclusión

La estadística Bayesiana comprende una serie de herramientas y de forma de interpretar la estadística diferente a la utilizada habitualmente cuando trabajamos con métodos de interpretación frecuentistas.

Esta forma de entender la estadística puede ayudarnos tanto a entender mejor cómo funciona la estadística general, a realizar mejores interpretaciones de los trabajos que leemos e incluso aplicar este tipo de estadística en nuestro día a día. Es por ello, que, tanto si se acaba aplicando en nuestra práctica laboral como si no, conozcamos algo sobre esta forma de interpretar la probabilidad que puede ser en algunas ocasiones más completa que la habitual y mirar a los datos no únicamente como cosas objetivas – que, además, pueden ser erróneas – sino globalmente con ayuda e interpretación activa por nuestra parte que puede ayudarnos a interpretar mejor los hallazgos, realizar inferencias y tener una mente más abierta a la hora de enfrentarnos a la parte estadística de nuestro trabajo, huyendo de la visión clásica de si un resultado es significativo o no o de la dependencia del p valor.

Este trabajo espera haber tenido una impresión positiva sobre el lector y haberlo animado, especialmente con el ejemplo práctico expuesto aquí, a seguir leyendo y estudiando sobre este tipo de estadística e interpretación de la probabilidad que puede aportarnos mucho. Se recomienda acudir a la sección *Referencias* de este trabajo donde estarán citados artículos y libros que ayudarán a entender mejor cómo funciona la estadística Bayesiana, tanto de forma más teórica y conceptual como de forma más práctica utilizando el lenguaje de programación R.

Por último, mencionar que todas las figuras han sido realizadas con el lenguaje de programación R mediante el uso de varios de los paquetes descritos en la diapositiva 19. Los valores de la única tabla de la presentación están adaptados de un artículo de S. Andraszewicz et al. (2015).

Todas las figuras son de uso libre mediante una licencia de cultura libre Reconocimiento 4.0 Internacional (CC BY4.0).



Referencias

1. Lambert B. *A student's guide to bayesian statistics*. London, England: SAGE Publications; 2018. <https://international.scholarvox.com/book/88869734>.
2. Kruschke JK, Liddell TM. Bayesian data analysis for newcomers. *Psychon Bull Rev.* 2018;25(1):155-177. <https://link.springer.com/article/10.3758/s13423-017-1272-1>. doi: 10.3758/s13423-017-1272-1.
3. Bürkner P. Advanced bayesian multilevel modeling with the R package brms. *The R journal.* 2018;10(1):395. https://explore.openaire.eu/search/result?id=doi_dedup___::40e00346279c87d56d861bfa7ab95154. doi: 10.32614/RJ-2018-017.
4. Bürkner P. Brms: An R package for bayesian multilevel models using stan. *Journal of statistical software.* 2017;80(1):1-28. https://explore.openaire.eu/search/result?id=doi_dedup___::959d79e6bce6b681ba96a71976f3cd20. doi: 10.18637/JSS.V080.I01.
5. Kruschke JK, Liddell TM. The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychon Bull Rev.* 2018;25(1):178-206. <https://link.springer.com/article/10.3758/s13423-016-1221-4>. doi: 10.3758/s13423-016-1221-4.
6. Kruschke JK. Bayesian estimation supersedes the T test. *Journal of experimental psychology. General.* 2013;142(2):573-603. <http://eric.ed.gov/ERICWebPortal/detail?accno=EJ1008485>. doi: 10.1037/a0029146.
7. Gelman A. *Bayesian data analysis*. Third edition ed. Boca Raton ; London ; New York: CRC Press; 2014. 10.1201/b16018.
8. Kruschke JK. *Doing bayesian data analysis*. edition 2 ed. Amsterdam ; Boston ; Heidelberg ; London ; New York ; Oxford ; Paris ; San Diego ; San Francisco ; Singapore ; Sydney ; Tokyo: Elsevier, AP; 2015. http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=027620605&sequence=000002&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA.
9. McElreath R. *Statistical rethinking*. Vol 122. Second edition ed. Boca Raton ; London ; New York: CRC Press; 2020. 10.1201/9780429029608.