



Inferencia multimodelo en ciencias sociales y ambientales

LUCAS A. GARIBALDI^{1,✉}; FRANCISCO J. ARISTIMUÑO²; FACUNDO J. ODDI¹ & FLORENCIA TIRIBELLI³

¹Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural (IRNAD), Sede Andina, Universidad Nacional de Río Negro (UNRN) y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), San Carlos de Bariloche, Río Negro, Argentina. ²Centro de Estudios en Ciencia, Tecnología, Cultura y Desarrollo (CITECDE), Sede Andina, Universidad Nacional de Río Negro (UNRN) y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), San Carlos de Bariloche, Río Negro, Argentina. ³Instituto de Investigaciones en Biodiversidad y Medioambiente (INIBIOMA), Universidad Nacional del Comahue (UNCo) y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), San Carlos de Bariloche, Río Negro, Argentina.

RESUMEN. Los profesionales de las ciencias sociales y ambientales debemos resolver problemas (contestar preguntas) a partir de la recolección y el análisis de datos. Habitualmente, todos enfrentamos dificultades similares; queremos tomar decisiones sobre una población (e.g., todos los árboles de una región), pero sólo contamos con información de una muestra (algunos árboles de esa región). Una herramienta fundamental en este proceso es plantear modelos de la población sobre la variable de interés (crecimiento los árboles en función de la edad y las condiciones climáticas) para luego utilizar sus predicciones en la toma de decisiones (turnos de corta de acuerdo a las condiciones climáticas). En esta ayuda didáctica discutimos cómo plantear, estimar y seleccionar modelos de una población a partir de los datos de una muestra. Dedicamos especial énfasis a proponer varios modelos (hipótesis) alternativos ante un mismo problema (e.g., distintas funciones del crecimiento arbóreo con la edad), los cuales son planteados antes de recolectar los datos e incluyen un modelo nulo (el crecimiento arbóreo no depende de la edad ni del clima). Los modelos nos indican cómo se deben recolectar los datos para un contraste válido (e.g., mediciones del crecimiento en árboles de edad distinta y en sitios con clima contrastante). Luego, el criterio de información de Akaike (AIC) permite ordenar los modelos según su parsimonia y seleccionar aquellos que mejor se ajusten a los datos (verosimilitud), y con menor número de parámetros (complejidad). A lo largo del texto introducimos las nociones básicas sobre la inferencia multimodelo y discutimos los errores más comunes en su uso. Proveemos ejemplos reales y hacemos disponibles los datos y los códigos de ejecución en el programa R, de acceso gratuito. Además de ser útil para los profesionales, esperamos que esta ayuda didáctica promueva la enseñanza de la inferencia multimodelo en los cursos de grado.

[Palabras clave: AIC, ajuste, Akaike, hipótesis, inferencia, modelo, parsimonia, predicción, valor P , verosimilitud]

ABSTRACT. Multimodel inference in social and environmental sciences. Professionals of the social and environmental sciences must solve problems (answer questions) based on data sampling and analyses. Commonly, all professionals face similar challenges: they need to take decisions on a population (e.g., all the trees of a region), but only have data from a sample (some trees of that region). A key tool in this process is to propose population models for the response variable (tree growth as a function of tree age and climatic conditions) and then use model predictions to take decisions (e.g., when to cut trees according to climatic conditions). In this paper we discuss how to propose, estimate, and select models of a population based on sampling data. We put special emphasis in proposing several alternative models (hypotheses) to solve one problem (e.g., different tree growth functions for age), which must be proposed before data sampling, including a null model (tree growth does not depend on tree age or climatic conditions). Models guide us on how data must be sampled for a valid contrast (growth measurements in trees of different age and under contrasting climates). Then, the Akaike information criterion (AIC) can be employed to sort the most parsimonious models, selecting those with the best goodness of fit (likelihood) and the lowest number of parameters (model complexity). Along the text, we introduce basic notions of multimodel inference and discuss common user mistakes. We provide real examples, and share their data and the analyses code in R, a free and open source software. In addition to be useful to professionals from different sciences, we expect our paper to promote the teaching of multimodel inference in graduate courses.

[Keywords: AIC, goodness of fit, Akaike, hypothesis, inference, model, parsimony, prediction, P value, likelihood]

INTRODUCCIÓN

Si contestamos correctamente las preguntas cuya respuesta es incierta, nuestro bienestar aumenta. Por ejemplo, ¿cuántos ratones hay en los domicilios de una ciudad y cuál es la relación con la transmisión de enfermedades al ser humano? Si bien no estaremos seguros de la respuesta, nuestro grado de certeza aumentará a medida que más y mejor información tengamos. En el ámbito profesional, la información se recolecta en forma de datos a partir de relevamientos (experimentos). En este ejemplo, un conteo de ratones por domicilio y de la presencia de enfermedades transmisibles al ser humano. Rápidamente nos enfrentamos con una serie de problemas; entre ellos, no podemos relevar todos los domicilios de una ciudad. Es decir, contamos con información incompleta (una muestra). Una vez decidido cuáles domicilios relevar y cómo relevarlos sin sesgo, la respuesta a nuestra pregunta se encuentra escondida en una tabla de datos que contiene información variable (los domicilios presentan distinto número de ratones y presencia de enfermedades). En este contexto, la estadística es una ciencia transversal a los distintos ámbitos profesionales, ya que nos brinda herramientas para recolectar un subconjunto de datos (diseño de estudios), resumir y presentar su información (estadística descriptiva) y arribar a conclusiones sobre la población (estadística inferencial). La estadística nos permite cuantificar la incertidumbre asociada a nuestras respuestas y, por lo tanto, las conclusiones a las que arribamos se sustentan en un marco de análisis sólido. En esta ayuda didáctica nos concentramos en cómo contestar preguntas a partir de datos mediante modelos estadísticos alternativos sobre la población (inferencia multimodelo).

La inferencia multimodelo se usa cada vez más frecuentemente en el ámbito científico y en el profesional. Sin embargo, es raro que se enseñe en los cursos de grado (Aho et al. 2014; Rosen 2016). Además, si bien existe gran cantidad de literatura sobre el tema, la mayoría se encuentra en textos especializados en inglés (e.g., Burnham et al. 2011). En este trabajo presentamos brevemente los fundamentos de la inferencia multimodelo y del criterio de información de Akaike (AIC por sus siglas en inglés), e incluimos la verosimilitud y el principio de parsimonia, además de contrastarlo con la inferencia clásica. Apuntamos a que el texto pueda ser comprendido por estudiantes de

grado que conocen los conceptos abordados habitualmente en un primer curso de estadística (como estadística descriptiva, probabilidad, distribuciones por muestreo, e inferencia), así como por profesionales de las ciencias sociales y ambientales. Para facilitar la comprensión brindamos un glosario de términos (Caja 1), que no pretende dar definiciones formales. Realizamos la introducción de los conceptos a partir de dos ejemplos, uno en el texto principal (producción de cereales) y otro en la Caja 2 (crecimiento de delfines). Para ambos ejemplos proveemos los datos y los códigos de los análisis en el ambiente de programación R (R Core Team 2016) como apéndices. Si bien trabajamos con datos reales, el énfasis está puesto en presentar de manera pedagógica la inferencia multimodelo, sin ser nuestra intención realizar avances ni dar recomendaciones sobre la temática abordada. Empecemos entonces con uno de los ejemplos.

PRODUCCIÓN DE CEREALES

Es de común acuerdo que en la producción de la gran mayoría de bienes consumidos por la sociedad participan el capital y el trabajo. Sin embargo, la forma en que interviene cada uno de ellos queda determinada por las condiciones productivas específicas de cada bien (en nuestro caso, la producción de cereales a nivel mundial). Estas formas se estudian mediante modelos que describen la cantidad de producto para diferentes valores de los predictores (en economía, estos modelos se conocen como funciones de producción). A lo largo de la historia se han propuesto varios modelos. Algunos de los más conocidos son el de “Cobb-Douglas” y el de “Elasticidad de Sustitución Constante (ESC)”. La explicación y el desarrollo de los mismos está fuera del alcance de este artículo y sólo diremos que el modelo de Cobb-Douglas (Cobb and Douglas 1928) es un caso particular del de ESC (Arrow et al. 1961; Chiang and Wainwright 2006).

En este ejemplo estudiaremos la producción de cereales para conocer su relación con el capital y el trabajo. Se apunta a estimar una relación funcional para los países que componen la población de referencia (ver restricciones más abajo). Los datos fueron extraídos de la FAO para el año 2006 (FAO 2014), e incluyen como variable “dependiente” la producción anual de cereales (en toneladas) y como variables “predictoras” la adición de stock de capital neto en desarrollo de la tierra, maquinaria y cultivos a precios

constantes (en millones de USD de 2005) y la población económicamente activa (PEA) en el agro (en miles de personas). Las variables predictoras son una aproximación a los factores productivos genuinos (capital y trabajo, respectivamente) y las utilizamos debido a que la información sobre lo invertido específicamente en la producción de cereales no está disponible. Limitamos el análisis (la población de referencia) a aquellos países que se especializan en la producción de cereales (al menos 20% de su PBI agropecuario proviene de los cereales), lo cual reduce la confusión que podría generar que los países destinen los recursos de capital y trabajo a otro tipo de cultivos (e.g., oleaginosas).

En esta ayuda didáctica nos proponemos analizar, mediante el uso de inferencia multimodelo, la bondad del modelo de Cobb-Douglas y el de ESC para explicar la producción de cereales a nivel global (Apéndice 1). Además, evaluaremos si la producción es mejor descrita por relaciones más simples (lineales) y también planteamos un modelo nulo (sin predictores) para tener un nivel de referencia. En resumen, analizaremos los siguientes modelos (hipótesis):

Modelo 0 - Nulo:

$$\text{cereales}_i \sim \text{Normal}(\mu; \sigma^2)$$

$$\mu = \beta_0$$

donde i indica el país (unidad experimental) y β_0 representa la media poblacional de producción de cereales. Este modelo indica que la producción de cereales sigue (ver símbolo “~”) una distribución normal (campana de Gauss) de probabilidades, compuesta por dos parámetros, una media ($\mu = \beta_0$) y una varianza (σ^2).

Modelo 1 - Lineal sin interacción entre los predictores:

$$\text{cereales}_i \sim \text{Normal}(\mu_i; \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 * \text{capital}_i + \beta_2 * \text{trabajo}_i$$

donde β_0 representa la producción media de cereales cuando el capital y el trabajo son nulos, β_1 es el efecto del capital y β_2 es el efecto del trabajo (β_1 y β_2 son conocidos como coeficientes parciales de regresión y reflejan el cambio medio en la variable dependiente por unidad de aumento en la variable independiente). Este modelo contiene muchas distribuciones normales (a diferencia del modelo nulo que

propone una sola distribución normal), las cuáles difieren en su media (μ_i , por eso ahora aparece el subíndice i al lado de μ) pero todas ellas presentan la misma varianza (σ^2). La media de cada distribución se obtiene a partir de una relación lineal con el capital y el trabajo.

Modelo 2 - Lineal con interacción entre los predictores:

$$\begin{aligned} \text{cereales}_i &\sim \text{Normal}(\mu_i; \sigma^2) \\ \mu_i &= \beta_0 + \beta_1 * \text{capital}_i + \beta_2 * \text{trabajo}_i + \beta_3 * \\ &\text{capital}_i * \text{trabajo}_i \end{aligned}$$

donde se agrega β_3 que representa el efecto de la interacción. Para más detalles sobre la interpretación de los coeficientes en los modelos lineales, se pueden consultar manuales de estadística de cursos de grado (e.g., Anderson et al. 2011).

Modelo 3 - Cobb-Douglas:

$$\text{cereales}_i \sim \text{Normal}(\mu_i; \sigma^2)$$

$$\mu_i = A (\text{capital}_i^{B1} + \text{trabajo}_i^{B2})$$

Modelo 4 - ESC:

$$\text{cereales}_i \sim \text{Normal}(\mu_i; \sigma^2)$$

$$\mu_i = A * [B * \text{capital}_i^{-R} + (1 - B) * \text{trabajo}_i^{-R}]^{-V/R}$$

donde A es un parámetro relacionado a la eficiencia tecnológica, B ($B1$ y $B2$ en el modelo 3) determina la participación relativa de los predictores, R se relaciona con la constante de elasticidad de sustitución y V mide el efecto de la escala de producción. No es necesario para avanzar en esta ayuda didáctica comprender en profundidad estos conceptos económicos, los cuáles pueden ser consultados en la literatura estándar sobre el tema (por ejemplo, (Cobb and Douglas 1928; Arrow et al. 1961; Chiang and Wainwright 2006). Lo clave aquí es que los parámetros de los modelos tienen un significado conceptual interesante para la problemática abordada y esto es algo a lo que siempre tenemos que apuntar a la hora de plantear cualquier modelo estadístico.

Como se puede observar, para simplificar nuestro estudio modelaremos únicamente la media de la producción de cereales a nivel mundial. Esto implica suponer que la varianza es la misma (homocedasticidad) para todas las observaciones. Otros supuestos importantes son el de independencia de las observaciones y la no existencia de correlación importante

entre variables predictoras (multicolinealidad). Finalmente, también suponemos que los residuos de cada uno de los modelos siguen una distribución normal. Entonces, el modelo 0 tiene dos parámetros (media y varianza) y está anidado en el modelo 1, de cuatro parámetros, ambos anidados dentro del modelo 2, de cinco parámetros. El modelo 3 cuenta con cuatro parámetros y se anida dentro del modelo 4, que posee cinco parámetros. A diferencia de los primeros, los modelos 3 y 4 no son lineales en sus parámetros.

¿QUÉ ES LA INFERENCIA ESTADÍSTICA? ¿POR QUÉ ES ÚTIL?

La estadística inferencial es la rama de la estadística que nos permite inferir ciertos aspectos de una población a partir de observar un subconjunto de la misma (la muestra). Dado que sólo trabajamos con una muestra, los resultados que obtenemos presentan incertidumbre, lo que quiere decir que no estaremos totalmente seguros de nuestra respuesta (Cox 2006). En nuestro ejemplo nos interesa explicar la variabilidad en la producción de cereales a nivel mundial (ver otro ejemplo en la Caja 2). Dado que necesitamos conocer en qué medida los predictores influyen en la variable dependiente, realizamos un experimento mensurativo (Hulbert 1984) en el que en cada país seleccionado recolectamos datos de la producción de cereales, del capital y del trabajo. Es interesante destacar que las variables capital y trabajo fueron elegidas en función de un marco teórico previo desde el cual postulamos nuestras hipótesis. Luego, para evaluar estas hipótesis, proponemos modelos estadísticos en los que se plantean las relaciones entre la producción de cereales y los predictores. Al llevar a cabo el análisis estadístico obtenemos una estimación de los parámetros de los modelos planteados, los cuales describen la influencia de cada predictor en la producción de cereales. Así, estamos “infiriendo” la magnitud de la influencia de los predictores en la producción de cereales a partir de relevar un subconjunto de países y, por lo tanto, tenemos incertidumbre. Si bien la conclusión sobre nuestro problema de interés (y, por lo tanto, las decisiones basadas en esta conclusión) estarán siempre a cargo del investigador, la inferencia estadística provee un marco de análisis que nos ayuda a tomar decisiones racionales.

DOS MARCOS CONCEPTUALES DE INFERENCIA ESTADÍSTICA

Hasta el siglo XIX, la ciencia estadística era descriptiva, y fue el desarrollo de la teoría de probabilidades lo que permitió el surgimiento de la estadística inferencial. Se suele tomar una correspondencia entre Pascal y Fermat en 1654 como el origen de la teoría de las probabilidades, aunque a su vez se admite que ésta no alcanzó un grado suficiente de sistematización hasta el seminal tratado de Laplace en 1774, los trabajos de Gauss y Laplace durante 1809-1811 y de Fisher en 1922 (Todhunter 1865; Hacking 2006; Hald 2007). Luego de sentadas las bases, la inferencia estadística clásica se desarrolla desde 1930 a partir de las ideas de Fisher, por un lado, y de Neyman y Pearson por el otro. Fisher establece una prueba de significancia con el argumento de que los datos proporcionan la evidencia en relación a una hipótesis “nula” establecida de antemano, usando como medida de evidencia la probabilidad de encontrar el estadístico observado (o un valor más extremo) si la hipótesis nula fuera verdadera (valor P). De esta manera, sugiere que hay evidencia de que la hipótesis nula es falsa en el caso de que el valor P sea menor que un cierto umbral (nivel de significancia). Por su lado, Neyman y Pearson indican la necesidad de formular dos hipótesis complementarias, una nula y una alternativa. Establecen el concepto de probabilidades de error, es decir, la probabilidad a priori de rechazar o no erróneamente la hipótesis nula (error de tipo 1 y 2, respectivamente) bajo una visión frecuentista de ensayos repetidos. Esta escuela enfatiza la elección entre dos hipótesis y establece el concepto de potencia de una prueba estadística (Hubbard and Bayarri 2003).

Si bien surgieron como escuelas de pensamientos enfrentadas (Hubbard and Bayarri 2003), la enseñanza estadística actual se basa en un híbrido de ambos enfoques, en el que se establece una hipótesis nula y una alternativa definida como el complemento de aquella, se fija una probabilidad admitida de equivocarse al rechazar la hipótesis nula (error de tipo 1) y se calcula el valor P en base a los datos muestrales, para luego concluir si se rechaza o no la hipótesis nula (Tabla 1 y Figura 1). Es importante, especialmente si no se rechaza la hipótesis nula, informar

Tabla 1. Algunas características de la inferencia clásica y multimodelo.**Table 1.** Some aspects of classical and multimodel inference.

	Inferencia clásica	Inferencia multimodelo
Contraste	Contrasta un estadístico obtenido de los datos muestrales con respecto a una hipótesis "nula" del valor de un parámetro en la población. Informa probabilidad de cometer un error de tipo 1 y de tipo 2.	Contrasta varios modelos anidados y no anidados de manera simultánea. Se intenta evitar el sesgo de "enamorarnos" de una sola hipótesis y ver en los datos información que la sustente, cuando en realidad en muchas situaciones hay más de una hipótesis factible de explicar los patrones observados en los datos.
Valor <i>P</i> y error de tipo 1	Indica la probabilidad de obtener un valor igual o más extremo al estadístico muestral si la hipótesis "nula" es verdadera. El valor <i>P</i> se compara con una probabilidad fijada a priori de cometer un error de tipo 1 (nivel de significancia).	No aplica.
Criterio de información de Akaike (AIC)	No aplica.	Se utiliza para comparar la parsimonia de los distintos modelos planteados. El contraste es relativo; nos indica cuál o cuáles modelos son los que tienen mayor parsimonia. Es deseable, por lo tanto, incorporar en el contraste un modelo nulo (sin predictores).
<i>r</i> ²	Se utiliza como índice de bondad de ajuste en aquellos casos en los que puede ser calculado, como los modelos lineales generales.	Otros índices como el AICc, QAIC, cAIC, BIC y DIC se utilizan con fines similares. Se utiliza como índice de bondad de ajuste en aquellos casos en los que puede ser calculado, como los modelos lineales generales. El <i>r</i> ² "ajustado" (el cual penaliza por complejidad del modelo) puede ser usado con fines similares a los del AIC.
Potencia = 1- <i>P</i> (error de tipo 2)	Se puede calcular dicha probabilidad al fijar una hipótesis alternativa de interés.	No aplica.
Tamaño muestral	Al aumentar el tamaño muestral la potencia aumenta, pero no influye sobre el nivel de significancia (y por lo tanto la probabilidad de cometer un error de tipo 1). Para una diferencia dada, a mayor tamaño muestral más probabilidad de rechazar la hipótesis nula.	Siempre es deseable tener mayor tamaño muestral, pero ello no influye de manera previsible sobre el orden (ranking) de los modelos a ser comparados. Cuando se emplea el AIC, todos los modelos a comparar deben tener el mismo tamaño muestral.

la potencia de la prueba (la cual debe ser definida por el profesional antes de recolectar los datos). Bajo este paradigma, se postula un modelo estadístico por vez, y se lo utiliza para determinar si las variables predictoras (cuantitativas o cualitativas) explican el patrón observado en la variable dependiente (Stanton-Geddes et al. 2014). Un aspecto interesante de este enfoque es que al momento de concluir sobre nuestra investigación tomaremos una decisión controlando el nivel de incertidumbre que enfrentamos. En el caso de rechazar la hipótesis nula, concluiremos, por ejemplo, que

el capital explica una parte significativa de la variación en la producción de cereales con una probabilidad de equivocarnos decidida por nosotros a priori (por ejemplo, 0.01 ó 0.05).

En este punto debemos señalar que si bien generalmente la hipótesis de trabajo es la influencia de una variable predictora (e.g., capital) sobre la variable dependiente (e.g., producción anual de cereales), la hipótesis estadística a evaluar (hipótesis nula) muchas veces expresa la ausencia de efecto, es decir, la situación donde el capital invertido en el agro

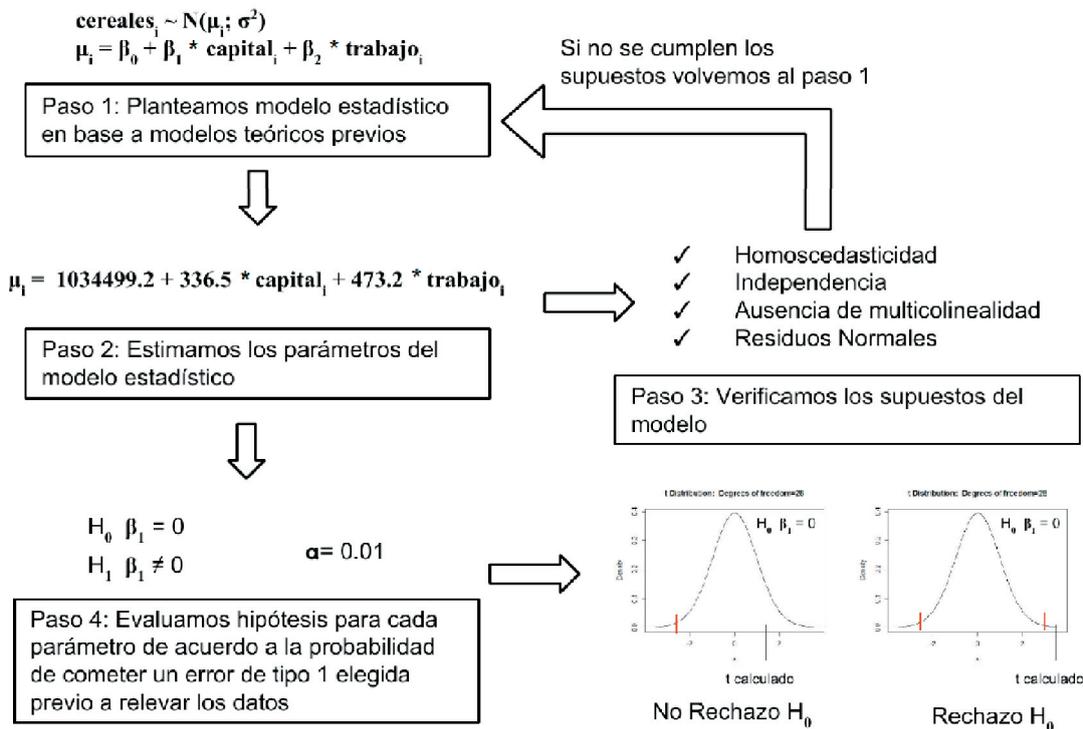


Figura 1. Breve descripción de contraste de hipótesis en inferencia clásica.

Figura 1. Brief description of hypothesis contrast in classical inference.

no explica la variabilidad en la producción anual de cereales (Tabla 1 y Figura 1). Sin embargo, en muchos casos la hipótesis nula es trivial y de poco interés para la disciplina. Si bien se puede plantear y evaluar otra hipótesis distinta a la trivial (e.g., el parámetro es igual a 1 ó, lo que es lo mismo, por cada unidad adicional de capital aumenta la producción media de cereales en una unidad), no es la aproximación que se suele emplear (Murtaugh 2014). Además, este contraste no nos indica la magnitud de la influencia de la variable predictora sobre la variable dependiente (e.g., cuánto aumenta la producción media de cereales por cada aumento en el capital); sólo nos permite concluir si esa variable tiene, o no, algún efecto (e.g., si el aumento en el capital afecta, o no, la producción media de cereales). De esta manera, lo máximo a lo que podemos aspirar es a falsar la hipótesis planteada o, en caso contrario, y en armonía con los principios del falsacionismo popperiano (Popper 1980), aceptarla hasta que se pruebe lo contrario (es decir, “no rechazarla”).

La inferencia clásica también ofrece otras herramientas valiosas que de alguna manera suplen las deficiencias del contraste de hipótesis. Por ejemplo, las estimaciones por medio de intervalos de confianza permiten

prestar más atención al tamaño del efecto, en lugar de encerrarse en la dicotomía, muchas veces irrelevante, de un contraste de hipótesis. Es por esto que hay cada vez más consenso en que el énfasis debe estar en la estimación y la evidencia, en lugar de en la decisión final entre rechazar y no rechazar (Laco Mazzone et al. 2016).

Ya en 1890, Chamberlin hablaba de las ventajas de trabajar con hipótesis múltiples en forma simultánea (Chamberlin 1890). Manejándonos con una única hipótesis podríamos a lo sumo verificar si existe información suficiente para rechazar la hipótesis de que la producción de cereales a nivel mundial sigue el modelo de Cobb-Douglas, o el de elasticidad de sustitución constante (ESC) o cualquier otro, pero siempre limitándonos al contraste de ese único modelo. Las herramientas estadísticas actuales, de la mano del desarrollo informático, nos permiten trabajar en simultáneo con muchos modelos para quedarnos con aquél o aquellos que tengan más evidencia en su favor (Tabla 1 y Figura 2). Dado que muchas veces hay más de una hipótesis (modelo) igualmente viable para explicar el fenómeno de interés, la inferencia multimodelo intenta evitar el sesgo de muchos profesionales de “enamorarse” de una sola hipótesis, y de encontrar información

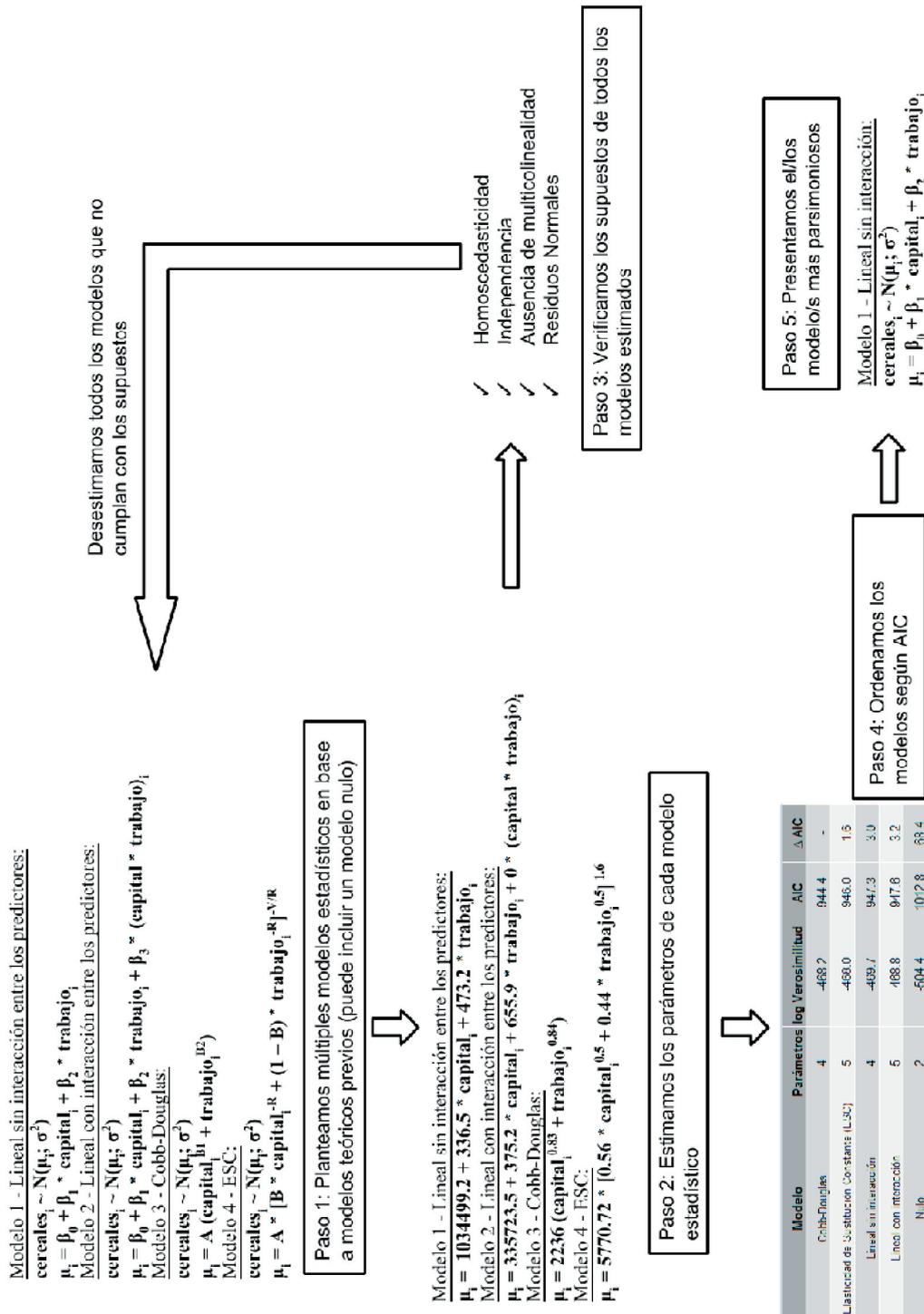


Figure 2. Breve descripción de inferencia multimodelo utilizando AIC. Ver Apéndices 1 y 2 para una aplicación de estos pasos de inferencia con datos reales en el programa R. Figure 2. Brief description of multimodel inference with AIC. See Appendices 1 and 2 for an application of these inferential steps on real data in R software.

en los datos que la sustenten (Rosen 2016). En este sentido, la inferencia multimodelo evita la regla de rechazo o no rechazo de hipótesis y ofrece una alternativa atractiva a los contrastes tradicionales basados en el análisis de la varianza (ANOVA) y la prueba de Tukey (Burnham et al. 2011; Stanton-Geddes et al. 2014).

El primer paso en inferencia multimodelo es hacer un esfuerzo sincero por encontrar un conjunto de hipótesis (modelos a priori) que, sobre la base de la experiencia y el conocimiento del investigador en su disciplina, sea apropiado para explicar el fenómeno de interés (Rosen 2016). Esta etapa es un momento central de la investigación y no debe ser tomada a la ligera (Tabla 1 y Figura 2). Al plantear estos modelos, el investigador construye el ámbito para contrastar su marco conceptual con la evidencia empírica. En el caso de la producción de cereales (ver otro ejemplo en la Caja 2), los modelos que se proponen a priori para estudiar la forma de la producción a nivel global no son arbitrarios, sino que surgen de marcos conceptuales de teoría económica, especialmente de teoría de la producción (Chiang and Wainwright 2006). Las formas funcionales representadas en cada uno de los modelos derivan de estudios previos que demostraron su pertinencia y en los que cada uno de sus parámetros admite una interpretación situada en el problema. Así, por ejemplo, en la función de Cobb Douglas (ver arriba, modelo 3) los parámetros B1 y B2 representan la elasticidad de producción del capital y del trabajo, es decir, cuál es la relación de sustitución entre los predictores para mantener un mismo nivel de producción.

La inferencia multimodelo permite a un investigador que parte de un conjunto de modelos a priori (un conjunto de hipótesis de trabajo relevantes y bien justificadas) calificarlos desde el más al menos apropiado (AIC y model weights), e incluso unificarlos en un mismo modelo que de alguna manera extraiga lo mejor de todos ellos (model averaging). Sin lugar a dudas, esto implica una gran ventaja, pero también un riesgo inminente. Quien use estas herramientas debe conocer sus fundamentos, potencialidades y limitaciones, y entender que no vienen a reemplazar el esfuerzo conceptual de formación y articulación de ideas, sino a complementarlo. A continuación, desarrollamos brevemente los fundamentos del enfoque de inferencia multimodelo (Tabla 1), para luego aplicarlo en la resolución del problema de la producción

de cereales. Nuestro artículo representa una introducción a la inferencia multimodelo basada en el AIC, en la que no desarrollaremos en detalle conceptos como model weights y model averaging. Tampoco desarrollaremos otros métodos de inferencia multimodelo como la validación cruzada (Arlot and Celisse 2010).

PARSIMONIA: UN BALANCE ENTRE BONDAD DE AJUSTE Y COMPLEJIDAD

El principio de parsimonia fue propuesto en el siglo XIV por William Ockham. Postula que si varias teorías son igualmente explicativas, la más sencilla es la más adecuada. Por un lado, es necesario contemplar en qué grado los modelos explican el proceso estudiado, es decir su bondad de ajuste. Por otro lado, desde el punto de vista de los modelos estadísticos, son más sencillos aquellos que tienen menos parámetros, y se considera que la linealidad es menos compleja que la no linealidad. Lo interesante de la aproximación multimodelo es que la parsimonia de varios modelos es comparada sin requerir que estos sean anidados (Caja 1) o que sólo uno de ellos sea el correcto (Aho et al. 2014). Esto contrasta con la inferencia según Neyman-Pearson y Fisher (inferencia “clásica”) mencionada anteriormente.

Un modelo parsimonioso debería basarse en un grupo de parámetros identificados por el investigador como conceptualmente relevantes, incluir —si fuera necesario— covariables e interacciones y ser lo más simple posible. Dicho de otro modo, tener la complejidad “justa” dado el objetivo para el cual fue planteado (Aho et al. 2014). Por ejemplo, supongamos que la producción de cereales sigue el modelo de Cobb Douglas. Por un lado, ajustar una relación lineal entre las variables sería un modelo demasiado simple, que no sería útil para predecir el patrón no lineal (problemas por mala bondad de ajuste). Por otro lado, ajustar un polinomio de séptimo grado mejoraría sensiblemente la bondad de ajuste, aunque sería demasiado complejo. Además, tampoco sería útil para predecir datos no observados, dado que introduce variabilidad errática y previene conclusiones generales debido a que sus parámetros no son conceptualmente relevantes. En general, siempre que crece el número de variables predictoras en un modelo (y por lo tanto el número de parámetros) aumenta la bondad de ajuste (entre modelos anidados). Aunque sea

por azar, el predictor agregado puede mejorar el ajuste (nunca reducirlo). Sin embargo, la paradoja es que si bien estos modelos son excelentes para predecir los datos observados en la muestra, son pésimos en predecir los valores que tomarán unidades experimentales no consideradas en la muestra con la que el modelo fue ajustado. En conclusión, no son deseables ni los modelos muy simples ni los extremadamente complejos. Es importante remarcar aquí que simple o complejo depende de los objetivos para los cuales fue planteado el modelo (Evans et al. 2013). Otro aspecto clave es que muchas veces se asume que modelos más simples son más generales en sus aplicaciones, pero esto no es necesariamente así (Evans et al. 2013).

AIC: UN ESTIMADOR DE PARSIMONIA

La pregunta que surge ahora es: ¿cómo hacemos para elegir entre los modelos propuestos? A principios de la década del setenta, Hirotugu Akaike conectó el método de máxima verosimilitud (Caja 1) con la teoría de la información para desarrollar el AIC. A continuación, presentamos de forma muy breve estos conceptos. Para más detalles, consultar la bibliografía citada (Burnham and Anderson 2004; Hobbs and Hilborn 2006; Murtaugh 2014).

Por un lado, la verosimilitud (L) es la probabilidad conjunta de los datos (x) dado un modelo con w parámetros:

$$L = g(x|\hat{w}) = P(Y_1) * P(Y_2) * \dots * P(Y_n)$$

donde Y_1, Y_2, \dots es el valor observado en cada unidad experimental (cada país en nuestro ejemplo), n es el tamaño muestral, P es probabilidad, y el acento circunflejo en \hat{w} significa que son los valores estimados de w . Como se asume que las unidades experimentales son independientes, la probabilidad conjunta resulta de multiplicar la probabilidad asociada al valor encontrado en cada unidad experimental. Cuanto más próximo a 1 sea L , los valores predichos por el modelo estarán más cerca de los datos. Por ejemplo, en el caso de una regresión lineal simple, cuanto mayor sea L , los puntos estarán más cerca de la línea recta que propone dicho modelo. Mediante el método de máxima verosimilitud se obtienen los valores de \hat{w} que maximizan L , es decir que se obtienen los valores estimados de los parámetros "ordenada al origen" y "pendiente" de la

recta de regresión lineal simple que hacen más probable observar los datos. Para encontrar este máximo, se utilizan derivadas sobre $g(x|w)$. Dado que es más fácil derivar sumas que multiplicaciones, y para reducir restricciones computacionales asociadas a multiplicar una serie larga de números cercanos a cero, generalmente se usa el $\log(L)$ en lugar de L ; recordemos que $\log(Y_1 * Y_2) = \log(Y_1) + \log(Y_2)$.

Por otro lado, la teoría de la información se ha centrado en el estudio de la información perdida cuando se aproxima un modelo mediante la utilización de otro modelo. Por ejemplo, ¿cuánta información se pierde cuando utilizamos un modelo normal para aproximar la distribución binomial? Estos estudios llevaron al criterio de información de Kullback-Liebert (K-L), que permite conocer la información perdida cuando se trata de aproximar datos generados por un modelo, digamos f , a través de otro, digamos g :

$$I_{K-L}(f,g) = E_f[\log f(x)] - E_f[\log g(x)]$$

donde $I_{K-L}(f,g)$ es el criterio de información de K-L cuando se aproxima el modelo f a través del modelo g , y E es esperanza (Caja 1). Este criterio de información es siempre positivo, es más grande cuanto peor es la aproximación de g a f , y sólo es 0 cuando g es idéntico a f (dando cuenta de que la aproximación es perfecta y que no hay pérdida de información).

Cuando ambos modelos están perfectamente especificados (se pueden escribir en fórmulas matemáticas), computar el coeficiente de información de K-L es sólo una cuestión de cálculos. El problema surge cuando queremos aplicar estos conceptos a la inferencia estadística, ya que, en ese caso, el modelo que queremos aproximar (f) es la realidad, para la cual, por más esfuerzo que hagamos, jamás tendremos una fórmula. Nuevamente, aquí vemos el problema fundamental de la inferencia: concluir sobre una población partiendo únicamente de una muestra.

Sin embargo, la inferencia multimodelo propone re-enfocar el problema aceptando que la realidad jamás va a ser descripta exhaustivamente por un modelo, y que no existe un modelo "verdadero". Observado desde esta perspectiva, el problema de la inferencia deja de ser un problema de distancias absolutas (de modelos propuestos a la realidad) para pasar a ser un problema de distancias relativas. El objetivo de la inferencia multimodelo no es

encontrar el modelo verdadero, sino el modelo más apropiado, reconociendo de antemano su falsedad en términos absolutos. Así, el primer término del criterio de información K-L ($E_f[\log f(x)]$) puede ser tratado como una constante cuyo valor no nos interesa porque es el mismo independientemente del modelo que estemos evaluando (la realidad es la misma, sin importar que modelo proponamos).

Entonces, el desafío en la búsqueda de un criterio de información relativo consistía en la estimación del segundo término de I_{K-L} :

$$I_r = E_f[\log g(x)]$$

Es interesante notar que para poder computar dicho término es necesario conocer los parámetros poblacionales (w) de $g(x)$. Sin embargo, como sólo contamos con una muestra, lo máximo que podemos conocer es \hat{w} , no w . Aquí no entraremos en detalles, pero un aporte fundamental de Akaike fue demostrar que el logaritmo de la verosimilitud ($\log(L)$) es un estimador sesgado (Caja 1) de I_r , ya que al aumentar el número de parámetros aumenta la verosimilitud. Akaike también demostró que ese sesgo tendía asintóticamente al número de parámetros (k) del modelo. De esta manera, si restamos el número de parámetros (k) obtenemos un estimador insesgado de I_r para muestras adecuadas y de gran tamaño:

$$\hat{I}_r = \log(L) - k$$

Esta solución pre-multiplicada por -2 es el criterio de información alcanzado por Akaike, más conocido como AIC.

$$AIC = -2 \log(L) + 2k$$

Así, el AIC de un modelo es proporcional a la probabilidad de los datos dado ese modelo (verosimilitud), penalizado por el número de parámetros (complejidad). Es aquí donde subyace la base del principio de parsimonia: explicar lo mejor posible el patrón observado en los datos con la mayor simpleza posible. Como está multiplicado por menos dos, valores más pequeños de AIC significan mayor parsimonia. El AIC es una medida relativa, que no nos dice que tan bueno es el modelo en sí mismo sino en comparación a otro. Es por ello que muchas veces resulta interesante que en la comparación incorporemos como referencia un modelo nulo (un modelo sin predictores). Es importante notar que L será más chica

cuantas más unidades experimentales tenga la muestra, dado que resulta de la multiplicación de una probabilidad (valores entre cero y uno) para cada unidad experimental. Por lo tanto, el valor del AIC depende del tamaño muestral y sólo es válido para comparar modelos con el mismo tamaño muestral.

¿INFERENCIA MULTIMODELO O PESCA DE MODELOS?

A nuestro criterio, estos métodos son usados correctamente cuando se proponen varias hipótesis (modelos) que se desprenden de la teoría, y resulta inadecuado llevar adelante una "pesca" de modelos. Esto último sería análogo a meter muchas variables predictoras en una "licuadora" y ver que combinación minimiza el AIC. Ello significa evaluar múltiples modelos sin una idea a priori que los sustente. La utilización de este último tipo de inferencia aumenta la probabilidad de encontrar relaciones por azar (bajo la inferencia clásica esto significa aumentar la probabilidad global de cometer un error de tipo I). Además, deberíamos recolectar otra muestra para validar un modelo que no expresa una hipótesis teórica. El lector interesado en estos temas puede abordar también la literatura sobre validaciones cruzadas en la inferencia multimodelo (Arlot and Celisse 2010; Pérez-Planells et al. 2015). Entonces, no es lo mismo proponer modelos estadísticos basados en uno o varios modelos conceptuales y luego contrastarlos con los datos, que proponer un modelo estadístico en base a los datos. El objetivo es explicar un fenómeno de la realidad (en nuestro caso la producción de cereales en los países del mundo) y no de los datos recogidos (la producción de cereales en los países de la muestra).

ALTERNATIVAS AL AIC

A lo largo de los años, se han propuesto varias alternativas al AIC. Entre ellas podemos listar al AICc, QAIC, cAIC, BIC y DIC (Grueber et al. 2011). A grandes rasgos, todos estos índices comparten el término $-2 * \ln(L)$ como estimador de bondad de ajuste, pero difieren en la manera en la que penalizan a los modelos por su complejidad (Grueber et al. 2011). No entraremos en detalle en cada una de estas alternativas, dado que el foco del artículo es presentar el uso de la inferencia multimodelo. Las recomendaciones brindadas aquí respecto de cómo llevar adelante la inferencia multimodelo y sus principios no

difieren de manera importante, ya sea que uno lleve adelante esta inferencia con AIC o, por ejemplo, con AICc (Tabla 1). En este texto hemos presentado el AIC dado que es el índice más utilizado (Aho et al. 2014). La discusión respecto las ventajas y desventajas relativas de cada índice, como por ejemplo ventajas de eficiencia versus consistencia en AIC versus BIC, respectivamente, pueden ser consultadas en la bibliografía citada (Grueber et al. 2011; Aho et al. 2014).

INFERENCIA MULTIMODELO EN LA PRODUCCIÓN DE CEREALES

Los supuestos subyacentes a todos los modelos son razonables (ver "Evaluación de supuestos" en el Apéndice 1) y, por lo tanto, son modelos válidos para ser evaluados (Figura 2). Si algún modelo no cumpliera con los supuestos, no sería apropiado y, por ende, no debería ser evaluado. La producción de cereales aumenta cuanto

mayor es el stock de capital y la población económicamente activa del agro (Figura 3), y todos los modelos que incorporan estos predictores presentan menor AIC respecto al modelo nulo (Tabla 2). Si embargo, estos modelos varían en el tipo de relación funcional con la producción de cereales (Figura 3). En particular, los datos muestran un mayor soporte hacia las relaciones no lineales de los modelos teóricos (Tabla 2). Entre éstos, si bien el modelo de Cobb-Douglas tiene un menor AIC, la diferencia respecto al modelo ESC no es concluyente (Tabla 2). Como es habitual en muchas situaciones de las ciencias sociales y ambientales, los datos indican que hay más de una hipótesis (modelo) factible.

¿VALOR *P* O AIC?

El valor *P* puede indicar qué tan incompatibles son los datos con un modelo estadístico específico (Murtaugh 2014; Laco Mazzone et al. 2016). Algunas críticas sobre el uso del valor

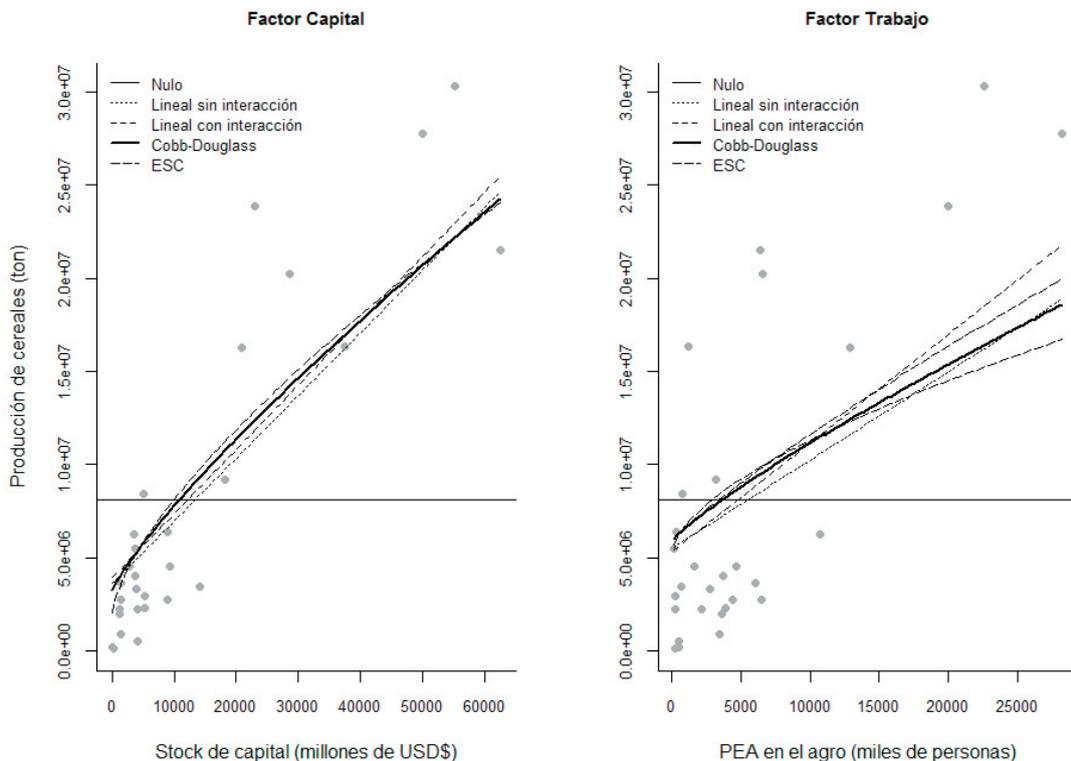


Figura 3. Producción de cereales en función del capital y trabajo (PEA=población económicamente activa) según los modelos ajustados (el modelo de Cobb-Douglas lo indicamos con trazo más grueso). En cada gráfico, las curvas son obtenidas fijando al otro predictor en su nivel medio.

Figure 3. Cereal production as a function of capital and labor (PEA=economic active population) according to the fitted models (the Cobb-Douglas model is indicated with a thick line). In each panel curves were obtained fixing the remaining predictor at the average.

Tabla 2. Inferencia multimodelo sobre la función de producción de cereales a nivel global. Los modelos son ordenados según su parsimonia, estimada a partir del criterio de información de Akaike (AIC), el cual es el resultado de la bondad de ajuste (log Verosimilitud) y complejidad (número de parámetros) de cada modelo.

Table 2. Multimodel inference on the cereal production function at the global level. Models are sorted according to their parsimony, based on the Akaike Information Criterion (AIC), which results from the goodness of fit (log likelihood) and its complexity (number of parameters).

Modelo	Parámetros	log Verosimilitud	AIC	Δ AIC
Cobb-Douglass	4	-468.2	944.4	-
Elasticidad de Sustitución Constante (ESC)	5	-468.0	946.0	1.6
Lineal sin interacción	4	-469.7	947.3	3.0
Lineal con interacción	5	-468.8	947.6	3.2
Nulo	2	-504.4	1012.8	68.4

P son una lógica de interpretación equivocada, aplicación tipo receta, confusión de la significancia estadística con la significancia en el campo de interés (ecología, economía, sociología, etc.), considerar al valor P como la probabilidad de que la hipótesis nula sea verdadera, considerar al valor P como la probabilidad de que el conjunto de datos haya sido generado por pura aleatoriedad, considerar al valor P como una medida del tamaño del efecto (Burnham and Anderson 2014; Laco Mazzone et al. 2016). Sin embargo, estos no son problemas intrínsecos del valor P , sino de una práctica estadística inadecuada por parte de los usuarios (Murtaugh 2014; Laco Mazzone et al. 2016). En efecto, la inferencia multimodelo no es inmune a las malas prácticas estadísticas, y en ocasiones la comparación de modelos mediante AIC es utilizada en forma análoga al contraste de hipótesis, concluyendo que dos modelos son diferentes cuando la diferencia de AIC es mayor, por ejemplo, a 2. En general, las conclusiones científicas y las decisiones empresariales o políticas no deben basarse sólo en si un valor P dado, o la diferencia en AIC entre dos modelos, supera un determinado umbral. Inferencias adecuadas requieren informes completos y transparentes.

Más allá de estas críticas al mal uso de la inferencia clásica, en la misma existe una limitación dado que no podemos ordenar diferentes hipótesis, a priori plausibles, bajo la evidencia dada por nuestros datos (Tabla 1). En este contexto sólo podemos condicionar los datos a las hipótesis nulas y luego decidir a favor o en contra de ellas. Por lo tanto, a esto se lo ve como una limitación bajo el paradigma de falsabilidad de las hipótesis de las ciencias empíricas (Popper 1980), ya que no contrastamos las hipótesis alternativas (que muchas veces son las de mayor interés), sino sólo las nulas (Burnham and Anderson 2014). En cambio, desde el

enfoque multimodelo se puede cristalizar la propuesta de Chamberlin (1890) de trabajar simultáneamente con hipótesis múltiples. Los problemas que deben afrontarse en las ciencias naturales como sociales requieren de un enfoque que permita el planteo simultáneo de numerosas alternativas ya que el resultado dicotómico de una única hipótesis (rechazo, no rechazo) dificulta explicar la complejidad de estos fenómenos. No obstante, y a diferencia del enfoque clásico, la inferencia multimodelo basada en AIC no permite cuantificar la incertidumbre asociada a la toma de decisiones mediante probabilidades de error y pruebas de potencia (Tabla 1).

CONCLUSIONES

Al igual que el resto las ciencias, la estadística se encuentra en un proceso continuo de cambio. Incluso el desarrollo de otras ciencias, como la informática, ha permitido el avance de la estadística. Uno de estos avances es la inferencia multimodelo realizada a través del AIC. Este índice puede ser usado para realizar inferencias (es decir, utilizar información de la muestra para concluir aspectos de la población) con modelos contrastantes (no anidados, con distribuciones estocásticas distintas, lineales vs. no lineales, etc.), los cuáles hasta hace relativamente poco tiempo debían ser analizados con programas estadísticos distintos, bajo nombres distintos y métodos contrastantes. Por lo tanto, el AIC abre las puertas a un uso más amplio de la inferencia, ya que permite comparar bajo una misma regla modelos muy distintos, que pueden ser explicaciones alternativas de un mismo fenómeno. Esta es la base de la inferencia multimodelo. Además, como en toda teoría, se valora el hecho de poder alcanzar mayor grado de generalidad (evaluar un amplio número de hipótesis o modelos) y mayor simpleza para abarcarla (principio de parsimonia).

Muchas veces hay más de una aproximación estadística adecuada para resolver un problema (Murtaugh 2014; Stanton-Geddes et al. 2014). En esta ayuda didáctica presentamos la inferencia multimodelo y usamos el AIC como ejemplo. Esta inferencia es muy interesante porque nos permite evaluar múltiples modelos (hipótesis) alternativos que explican (y predicen) un fenómeno de interés (sin requerir que sólo un modelo sea el correcto). Este aspecto, además de la posibilidad de trabajar con modelos muy distintos según fue explicado en el párrafo anterior, son ventajas claras con respecto al contraste de hipótesis nula basado en valor P . Sin embargo, este último método puede ser empleado de manera correcta en muchas situaciones y permite cuantificar cuán probable es equivocarnos al tomar una decisión (error de tipo 1 ó de tipo 2) (Murtaugh 2014; Stanton-Geddes et al. 2014). Cualquier aproximación de inferencia estadística no está libre de ser utilizada de forma incorrecta por los profesionales. Por ejemplo, así como no sería adecuado basar decisiones solamente evaluando si el valor P es menor o no a un límite arbitrario, comúnmente de 0.05 (Laco Mazzone et al. 2016), tampoco sería adecuado utilizar una diferencia de $AIC=2$ en el mismo contexto. La resolución de problemas en base a datos, y el avance científico, no dependen solamente de la elección de un método de inferencia, sino de muchos otros aspectos irremplazables como, por ejemplo, un sólido marco conceptual, una recolección de datos adecuada tanto en cantidad como en calidad, una discusión detallada enfatizando aspectos como el tamaño de efecto y una presentación gráfica clara.

MAPA CONCEPTUAL

Los mapas conceptuales nos ayudan a comprender (y a recordar) de una manera sintética las relaciones entre conceptos. Proponemos que el lector realice un mapa (diagrama) conceptual con las palabras clave de este artículo (conceptos) para profundizar la comprensión de las ideas aquí vertidas. En el mapa una con flechas (conectores) las palabras clave relacionadas. Enumere cada una de las flechas, debajo del mapa liste los números y al lado de cada número explique brevemente la relación entre los conceptos. Puede agregar hasta dos palabras clave si las necesita. No deben quedar conceptos aislados y puede haber más de una flecha que llegue o salga de cada concepto. No utilice más de una carilla tamaño A4 en su mapa y respuesta. Es importante no caer en la trampa de enumerar definiciones sueltas, relacione los conceptos entre sí e interprete.

AGRADECIMIENTOS. Marcelo Aizen, Sebastián Aguiar, Agustina Bazalote Oliver, Aliosha Behnisch, Gerardo de la Vega, Agustina di Virgilio, Florencia Fasani, María Laura Halladjian, Diego Nabaes Jordan, Amaru Magnin, Fiorella Laco Mazzone, Agustín Saez, Andrés Tálamo, Cristian Daniel Torres y Brenda Winkelman proveyeron comentarios y discusiones muy enriquecedoras. Agradecemos también el financiamiento de la Universidad Nacional de Río Negro en investigación (PI 40-B-399) y extensión (Resolución Rectoral 315/16), así como de la Agencia Nacional de Promoción Científica y Tecnológica (FONCYT, PICT 2013-1079).

REFERENCIAS

- Aho, K., D. Derryberry, and T. Peterson. 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* **95**:631-636.
- Anderson, D. R., D. J. Sweeney, and T. A. Williams. 2011. *Estadística para negocios y economía*. 11a. ed. Cengage Learning, Distrito Federal, México.
- Arlot, S., and A. Celisse. 2010. A survey of cross-validation procedures for model selection. *Statistics Surveys* **4**:40-79.
- Arrow, K. J., H. B. Chenery, B. S. Minhas, and R. M. Solow. 1961. Capital-Labor Substitution and Economic Efficiency. *The Review of Economics and Statistics* **43**:225-250.
- Brown, J. E., H. A. Fitzhugh, and T. C. Cartwright. 1976. A comparison of nonlinear models for describing weight-age relationships in cattle. *Journal of Animal Science* **42**:810-818.
- Burnham, K. P., and D. R. Anderson. 2014. P values are only an index to evidence: 20th- vs. 21st-century statistical science. *Ecology* **95**:627-630.
- Burnham, K. P., D. R. Anderson, and K. P. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* **65**:23-35.
- Burnham, K. P., and R. P. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research* **33**:261-304.

- Casas, G. A., D. Rodríguez, and G. Afanador Téllez. 2010. Propiedades matemáticas del modelo de Gompertz y su aplicación al crecimiento de los cerdos. *Revista Colombiana de Ciencias Pecuarias* **23**:349-358.
- Chamberlin, T. C. 1890. The method of multiple working hypotheses. *Science (New York, N.Y.)* **15**:92-96.
- Chiang, A. C., and K. Wainwright. 2006. *Métodos fundamentales de economía matemática*. Cuarta edi. McGraw-Hill Interamericana, México, D. F.
- Cobb, C. W., and P. H. Douglas. 1928. A theory of production. *The American Economic Review* **18**:139-165.
- Cox, D. R. 2006. *Principles of statistical inference*. Cambridge University Press, Cambridge, UK.
- Crespo, E., A. C. Schiavini, G. Pérez Macri, L. Reyes, and S. L. Dans. 1994. Estudios sobre la determinación de edad en mamíferos marinos del Atlántico Sudoccidental. Pages 31-55 in J. A. Oporto (ed.). *Anales de la 4a Reunión de Trabajo de Especialistas en Mamíferos Acuáticos de América del Sur*. Valdivia, Chile.
- Evans, M. R., V. Grimm, K. Johst, T. Knuuttila, R. de Langhe, C. M. Lessells, M. Merz, M. A. O'Malley, S. H. Orzack, M. Weisberg, D. J. Wilkinson, O. Wolkenhauer, and T. G. Benton. 2013. Do simple models lead to generality in ecology? *Trends in Ecology and Evolution* **28**:578-583.
- FAO. 2014. FAOSTAT. <http://faostat.fao.org/site/377/default.aspx#ancor>.
- Fernández, S., and A. A. Hohn. 1998. Age, growth, and calving season of bottlenose dolphins, *Tursiops truncatus* off coastal Texas. *Fishery Bulletin* **96**:357-365.
- Gruuber, C. E., S. Nakagawa, R. J. Laws, and I. G. Jamieson. 2011. Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology* **24**:699-711.
- Hacking, I. 2006. *The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference*. Second Edi. Cambridge University Press, Cambridge, UK.
- Hald, A. 2007. *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. Springer-Verlag, New York, USA.
- Hobbs, N. T., and R. Hilborn. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self-teaching. *Ecological Applications* **16**:5-19.
- Hubbard, R., and M. J. Bayarri. 2003. Confusion over measures of evidence (p 's) versus errors (α 's) in classical statistical testing. *The American Statistician* **57**:171-178.
- Hulbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**: 187-211.
- Laco Mazzone, F., M. Grampa, M. Goldenberg, F. Aristimuño, F. Oddi, and L. A. Garibaldi. 2016. Declaración de la Asociación de Estadística Americana sobre la significancia estadística y los valores P (editado por Ronald L. Wasserstein). *The American Statistician* **70**:Online discussion.
- Murtaugh, P. A. 2014. In defense of P values. *Ecology* **95**:611-617.
- Nobre, P. R. C., I. Miszta, S. Tsuruta, J. K. Bertrand, L. O. C. Silva, and P. S. Lopes. 2003. Analyses of growth curves of Nellore cattle by multiple-trait and random regression models. *Journal of Animal Science* **81**:918-926.
- Oltjen, J. W., A. C. Bywater, R. L. Baldwin, and W. N. Garrett. 1986. Development of a dynamic model of beef cattle growth and composition. *Journal of Animal Science* **62**:86-97.
- Pérez-Planells, L., J. Delegido, J. P. Rivera-Caicedo, and J. Verrelst. 2015. Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos. *Revista de Teledetección* **44**:55-65.
- Popper, K. R. 1980. La lógica de la investigación científica. Page Estructura y Función. El porvenir de la Ciencia. Editorial Tecnos S. A. Editorial Tecnos, Madrid, España.
- R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ratkowsky, D. A. 1983. *Nonlinear regression modeling: a unified practical approach*. Marcel Dekker Inc, New York, USA.
- Rosen, J. 2016. A forest of hypotheses. *Nature* **536**:239-241.
- Stanton-Geddes, J., C. G. De Freitas, and C. De Sales Dambros. 2014. In defense of P values: comment on the statistical methods actually used by ecologists. *Ecology* **95**:637-642.
- Todhunter, I. 1865. *A history of the mathematical theory of probability from the time of Pascal to that of Laplace*. Macmilland and Co, Cambridge and London.

CAJA 1. GLOSARIO

Distribución de probabilidades. Es una función matemática que asigna la probabilidad de que la variable (X) tome un valor determinado ($X=x_i$). Por lo tanto, si una distribución de probabilidades es dada por $Y=f(X)$, el eje Y representa probabilidades y el eje X el valor de la variable. El eje X es definido dentro del rango de valores que la variable puede tomar y se denomina espacio muestral (por ejemplo, algunas variables sólo pueden tomar valores positivos, otras se encuentran entre cero y uno). La distribución de probabilidades puede entenderse como una distribución de frecuencias teórica ya que modela cuán probable (o frecuente) es que un evento ocurra. En términos formales, es necesario distinguir entre la distribución de probabilidades de variables discretas y continuas. En estas últimas, la variable puede tomar infinitos valores dentro de un intervalo y por lo tanto la probabilidad de que X tome un valor exacto es cero. Es así que para variables continuas se habla de densidad de probabilidades.

Error de tipo 1. Rechazar la hipótesis nula siendo que es verdadera.

Error de tipo 2. No rechazar la hipótesis nula siendo que es falsa.

Esperanza. En estadística el concepto de esperanza (o valor esperado) hace referencia al valor medio de una variable aleatoria (X) que obtendríamos al repetir el experimento infinitas veces. Matemáticamente, resulta de sumar las multiplicaciones de los posibles valores de la variable ($X=x_i$) por la probabilidad (P) de obtener estos valores: $E(X) = \sum X = x_i * P(X=x_i)$.

Estadístico. A diferencia del parámetro, el estadístico es función de los datos de la muestra y por lo tanto depende de (es decir, varía con) la misma. Esto conlleva a que el estadístico tenga asociada una distribución por muestreo que, conceptualmente, surge de tomar infinitas muestras de un tamaño dado (cuanto mayor es el tamaño muestral, la variabilidad de esta distribución generalmente es menor). Como ejemplo podemos citar la media muestral, la varianza muestral, la proporción muestral y la pendiente muestral entre dos variables.

Estimación. Al calcular el estadístico en la muestra "estimamos" el parámetro en la población. La estimación puede ser puntual o mediante intervalos. La primera es el valor puntual obtenido en la muestra y no tiene asociado ningún nivel de incertidumbre. Dado que el estadístico proviene de una distribución (por muestreo) y que de acuerdo al Teorema Central del Límite es posible conocer la forma de esta distribución, podemos cuantificar su variabilidad y fijar una probabilidad a priori (confianza) de que el parámetro poblacional sea abarcado por cierto intervalo, el cual se denomina intervalo de confianza. En este caso, la estimación tiene asociada un nivel de incertidumbre.

Estimador sesgado. El estimador se considera sesgado cuando su esperanza difiere del valor del parámetro que estima.

Experimento. Procedimiento mediante el que se obtienen los valores de la(s) variable(s) de interés.

Modelos anidados. Un modelo está anidado dentro de otro cuando el de menor complejidad puede ser obtenido desde el más complejo al fijar, en este último, uno o más de sus parámetros a un determinado valor (en general cero).

Modelo estadístico. Está integrado por una componente estocástica que modela la variabilidad de los datos respecto al valor que predice el modelo, por ejemplo, a través de una distribución normal, y una componente determinística, la cual modela uno o varios parámetros en función de otras variables o predictores. Por ejemplo, el modelo de regresión lineal simple modela la media (parámetro) de una distribución normal (componente estocástica) a partir de una recta (componente determinística) entre X (predictor) e Y (variable respuesta). Cuando la componente determinística no tiene predictores asociados se dice que el modelo es nulo.

Muestra. Subconjunto de elementos (unidades experimentales) de la población desde la cual se pretende realizar inferencias. Para que las inferencias sean válidas la muestra debe ser representativa de la población bajo estudio.

Parámetro. Constante que caracteriza o representa a la población en algún aspecto. Su valor es función de los datos de la población; como ejemplo podemos citar la media poblacional, la varianza poblacional, la proporción poblacional, la pendiente poblacional entre dos variables. No es posible conocer el parámetro poblacional a menos que relevemos el total de la población.

Parsimonia. En el contexto de la teoría de la información, la parsimonia hace referencia a la situación de compromiso entre lo explicado por un modelo (bondad de ajuste) y la complejidad del mismo (número de parámetros). Modelos con muy buena bondad de ajuste, pero con un excesivo número de parámetros, no son parsimoniosos.

Potencia estadística. Probabilidad de rechazar la hipótesis nula siendo que es falsa dada una hipótesis alternativa particular como verdadera.

Unidad experimental. Entidad física (objeto, porción del espacio o segmento temporal) sobre la cual se realiza el experimento. Aquí seguimos la definición de Hulbert (1984), que habla de experimentos manipulativos y mensurativos y utiliza el concepto de unidad experimental para ambos tipos de experimentos.

Teoría de la información. La teoría de la información es una corriente teórica iniciada en la década de 1940. Es una rama de la teoría matemática y de las ciencias de la computación que estudia la información y todo lo relacionado con ella: canales, compresión de datos y criptografía, entre otros.

Valor P . Es la probabilidad, bajo un modelo estadístico dado, que un estadístico de los datos (por ejemplo, la diferencia de la media muestral entre dos grupos) sea igual que, o más extremo que, su valor observado. En este caso, el modelo estadístico hace referencia a la distribución por muestreo esperada bajo la hipótesis nula.

Verosimilitud. La verosimilitud (L) es la probabilidad (P) de observar los datos bajo un modelo estadístico determinado. Si tenemos un conjunto X de datos y un modelo g cuyos parámetros los representamos con w , entonces la verosimilitud es la probabilidad de observar $X=x$ dado que $w=w$: $L=g(X=x|w=w)$. Por ejemplo, si al lanzar tres veces la moneda obtuvimos tres caras (x =cara, cara, cara), la verosimilitud de ese resultado, si la probabilidad individual de "cara" fuera 0.5 ($w=0.5$), sería igual a $P(\text{cara}|w=0.5) * P(\text{cara}|w=0.5) * P(\text{cara}|w=0.5)$. El desarrollo formal de este ejemplo simple surge de calcular la probabilidad de obtener tres éxitos en tres intentos bajo un modelo de distribución de probabilidades binomial en el que la probabilidad individual de éxito es igual a 0.5.

CAJA 2. MODELOS DE CRECIMIENTO

En las ciencias agronómicas y biológicas es interesante estudiar cómo varía el tamaño de los individuos en función de su edad. Con este fin se ajustan modelos (curvas) que sirven para describir y predecir el crecimiento de los animales (Ratkowsky 1983). Por ejemplo, estos modelos tienen aplicaciones en la producción animal para saber a qué edad los individuos alcanzan el peso adecuado para ser comercializables. Una apropiada descripción de la dinámica de crecimiento de los animales comercializables es útil también para comparar tasas de crecimiento y predecir el comportamiento productivo de diferentes lotes en una empresa. Así, estos modelos permiten describir características importantes, tales como la precocidad, la ganancia diaria, el peso adulto y el intervalo de tiempo entre el nacimiento y la madurez (Brown et al. 1976; Oltjen et al. 1986; Nobre et al. 2003; Casas et al. 2010) y son, por lo tanto, herramientas clave para tomar decisiones financieras y de manejo.

El modelo de crecimiento más conocido es el desarrollado por el biólogo austriaco Ludwig von Bertalanffy a principios de la segunda mitad del siglo XX, quien realizó este modelo para describir el tamaño de los individuos de una población de peces en función de su edad. Este modelo supone que el crecimiento de los individuos es rápido al principio y posteriormente va disminuyendo hasta ser nulo (es decir que los individuos ya no pueden crecer más). Ludwig von Bertalanffy probó que si denominamos $Y(t)$ al tamaño de un individuo en función de su edad (t), a es el tamaño máximo que alcanzan los individuos de dicha población y b es la tasa de crecimiento:

$$Y(t) = a * (1 - e^{-(b * t)})$$

Otro modelo de crecimiento muy utilizado es el de Benjamín Gompertz, inicialmente propuesto en 1825 para describir la mortalidad humana en edades adultas. Muchas compañías lo usan en la actualidad para el cálculo de los costos de los seguros de vida. Este modelo supone que el crecimiento de los individuos disminuye de forma exponencial con el tiempo:

$$Y(t) = a * e^{-e^{-(c * (b * t))}}$$

en donde a es el tamaño máximo alcanzado por los individuos, b la tasa de crecimiento y c es una constante de proporcionalidad.

El modelo de crecimiento logístico también es uno de los más usados. Este tiene varias expresiones y cada usuario elige aquella que considera más útil para el problema que aborda. El modelo logístico es similar a los otros dos, supone un crecimiento inicial exponencial que va desacelerando con el tiempo hasta alcanzar un máximo, que será la talla del animal por el resto de su vida adulta:

$$Y(t) = a / (1 + e^{(c * (b * t))})$$

en donde a es el tamaño máximo, b es la tasa de crecimiento y c es una constante de proporcionalidad.

En este caso vamos a estudiar como varía el tamaño de machos adultos del delfín *Tursiops truncatus* en función de su edad. Conocido como delfín mular o delfín nariz de botella, *T. Truncatus* se distribuye según la CMS (Convention on Migratory Species) a lo largo de las costas tropicales y templadas del mundo. En Argentina se encuentra a lo largo de toda la costa Atlántica, incluso hasta Tierra del Fuego. Por ser carismático y muy inteligente se lo usó en varias películas como, por ejemplo, Flipper (1963), dirigida por James B. Clark. En este ejemplo utilizamos datos del trabajo realizado por Stephanie Fernandez y Aleta A. Hohn en las costas de Texas (Estados Unidos) (Fernandez and Hohn 1998). Los autores estimaron la edad y el largo de 78 delfines macho que vararon en la costa de Texas, además de otras variables de interés como, por ejemplo, el estado reproductivo (Apéndice 2). La edad de los delfines varados la estimaron contando las bandas de crecimiento de sus dientes, método muy utilizado en estudios de dinámica poblacional de cetáceos (Crespo et al. 1994). Los datos necesarios para ajustar los modelos propuestos son, entonces: a) la edad de los individuos medida en años como predictor (variable cuantitativa continua ya que para obtenerla se contaron las bandas de crecimiento de los dientes de estos cetáceos varias veces y se promedia), y b) su tamaño medido en centímetros como variable dependiente (Apéndice 2).

Preguntas.

- 1) Indicar unidad experimental, muestra y población.
- 2) Realice un gráfico del tamaño de los individuos en función de su edad. Rotule claramente los ejes del gráfico con sus unidades. Interprete en términos del problema.
- 3) Estime el modelo de von Bertalanffy, interprete los valores estimados de sus parámetros en términos del problema e indique las unidades de los parámetros (ayuda: de valores iniciales en R para que la función logre convergencia). Agregue al gráfico del punto (2) los valores predichos por el modelo de von Bertalanffy. Interprete en términos del problema.
- 4) ¿Cuál es la utilidad del error estándar asociado a la estimación de los parámetros? Explique utilizando el valor de algún error estándar del modelo de von Bertalanffy. Indique sus unidades. ¿Qué es una distribución por muestreo?
- 5) Compare la parsimonia de los tres modelos planteados (ayuda: de valores iniciales en R para que la función logre convergencia). Concluya. ¿En qué consiste el principio de parsimonia? Escriba (puede ser en papel aparte) el modelo con mayor parsimonia.