

Más allá de la cronología: uso de la inferencia Bayesiana para evaluar hipótesis en arqueología

Erik Otárola-Castillo (eoc@purdue.edu) , Melissa G. Torquato, Jesse Wolfhagen , Matthew E. Hill Jr. , y Caitlin E. Buck 

Erik Otárola-Castillo, Melissa G. Torquato y Jesse Wolfhagen Department of Anthropology, Purdue University, West Lafayette, Indiana, EE.UU. (eoc@purdue.edu, autor correspondiente)

Matthew E. Hill, Jr. Department of Anthropology, University of Iowa, Iowa City, Iowa, EE.UU.

Caitlin E. Buck School of Mathematics and Statistics, University of Sheffield, South Yorkshire, Reino Unido

RESUMEN

Los arqueólogos utilizan con frecuencia distribuciones de probabilidad y la prueba de significancia de la hipótesis nula (NHST por sus siglas en inglés) para evaluar cuán bien se alinean los datos de estudios, excavaciones o experimentos con sus hipótesis sobre el pasado. La inferencia bayesiana se usa cada vez más como alternativa a la NHST y, en arqueología, se aplica generalmente en la estimación de fechas de radiocarbono y la construcción de cronologías. Este artículo demuestra que las estadísticas bayesianas tienen aplicaciones más amplias. Comienza contrastando los marcos estadísticos de la NHST y bayesiano, antes de introducir y aplicar el teorema de Bayes. Con el fin de guiar al lector a través de un análisis bayesiano elemental paso a paso, este artículo utiliza un conjunto de fauna arqueológica de un sitio ficticio. Luego el ejemplo simulado se amplía para demostrar cómo se pueden aplicar los análisis bayesianos a datos con una variedad de propiedades, incorporando formalmente el conocimiento previo de los expertos en el proceso de evaluación de hipótesis.

Palabras clave: Estadística bayesiana, modelos de probabilidad frecuencial, contraste de hipótesis, arqueostatística.

INTRODUCCIÓN

Muchos arqueólogos están familiarizados con las estadísticas bayesianas en el contexto de la calibración de fechas por radiocarbono y la construcción de cronologías. Sin embargo, el marco bayesiano tiene aplicaciones más amplias que exceden a la datación y la cronología y que merecen ser consideradas por ellos. Por ejemplo, muchos investigadores de las ciencias naturales y sociales utilizan las estadísticas bayesianas para evaluar cuán bien se alinean los datos observacionales o experimentales con sus hipótesis. Por lo general, este uso de la inferencia bayesiana no se ha aplicado a la arqueología. Utilizando un ejemplo zooarqueológico ficticio, este artículo proporciona una explicación directa de la inferencia bayesiana y la compara con la prueba de significación de hipótesis nula (NHST por sus siglas en inglés) más convencional. Aunque algunos han descrito y revisado previamente la aplicación de estos conceptos en otros lugares (por ejemplo, Buck y Meson 2015; Buck et al. 1996; Otárola-Castillo y Torquato 2018; Otárola-Castillo et al. 2022; Wolfhagen 2019, 2020), este artículo está centrado en presentar ejemplos reproducibles paso a paso del marco bayesiano con la finalidad de evaluar y discernir entre hipótesis contrapuestas. El código Rmarkdown para reproducir todos los materiales presentados aquí está disponible en un repositorio de OpenScience Framework: <https://osf.io/23bdt/>. Además, el manuscrito original — escrito en

inglés — está disponible con los materiales complementarios de este artículo. Asimismo, los lectores en inglés pueden encontrar un código reproducible aquí: <https://osf.io/54f62/>.

INCERTIDUMBRE Y PROBABILIDAD EN APLICACIONES ARQUEOLÓGICAS

Todos los datos son inciertos. Las mediciones y observaciones no son exactas y sus valores resultantes son variablemente imprecisos. Los arqueólogos suelen utilizar cantidades estadísticas como la varianza, la desviación estándar y el error estándar, que están basados en la teoría de la probabilidad, para describir esta incertidumbre. En su trabajo de campo y laboratorio, ellos emplean equipos basados en descripciones probabilísticas de incertidumbre. Por ejemplo, el fabricante de estaciones totales —ampliamente utilizadas para mapear sitios arqueológicos— ha establecido precisiones de 2 mm más 2 mm adicionales por kilómetro, generalmente en el nivel de desviación estándar de 1σ (por ejemplo, Leica TS16). Este es un ejemplo de un concepto de probabilidad utilizado para medir la incertidumbre "aleatoria". En este caso, suponiendo una distribución de probabilidad "normal" para el error de medición (aunque el fabricante no lo especifica), los arqueólogos deben esperar que el 68 % de las ubicaciones

Por favor cite este trabajo como:

Otárola-Castillo, Erik, Torquato, Melissa G., Wolfhagen, Jesse, Hill, Matthew E., Jr., & Buck, Caitlin E.

2022 Más allá de la cronología: uso de la inferencia Bayesiana para evaluar hipótesis en arqueología. *Advances in Archaeological Practice*.

Traducido por Erik Otárola-Castillo. <https://doi.org/10.1017/aap.2022.10>.

de los artefactos cartografiados por este instrumento tengan un error de hasta ± 2 mm, más el error relacionado con el incremento de la distancia (y el error debido a las condiciones atmosféricas, la estabilidad del instrumento, etc.; Walker y Awange 2020). De forma similar, la hoja de especificaciones del fabricante para una báscula digital portátil típica de Ohaus (Scout STX2202) afirma medir hasta 2200 g, con un error de ± 0.02 g (1σ). Al igual que las estaciones totales, si asumimos un modelo de error normal, esto quiere decir que el fabricante certifica que el 68 % de todas las lecturas estarán dentro de ± 0.02 g de la lectura real en circunstancias ideales.

Análogamente, después de una cuidadosa recopilación y análisis de datos, los arqueólogos también aplican el concepto de probabilidad para probar sus hipótesis. Estas son declaraciones formales que ofrecen explicaciones plausibles de los patrones observados sobre las personas o su entorno en el pasado. Al igual que las afirmaciones sobre las mediciones de instrumentos de campo y laboratorio, estas hipótesis y sus predicciones también poseen cierto grado de incertidumbre debido a la observación o conocimiento incompletos. Para cuantificar de manera formal la incertidumbre sobre los datos y las hipótesis, los arqueólogos suelen confiar en modelos de probabilidad específicos o funciones de probabilidad (es decir, ecuaciones). Las entradas de una función de probabilidad son valores observados o hipotéticos, y los resultados son sus probabilidades que van de 0 a 1, es decir, de menos probable a más probable. Los arqueólogos utilizan este sistema probabilístico para probar sus hipótesis y describir el grado de incertidumbre con el que ellas dan cuenta de las observaciones actuales y futuras probables. El uso de un planteamiento probabilístico les ofrece una herramienta poderosa y sistemática que posibilita la interpretación de datos y la evaluación de hipótesis.

A continuación proporcionamos una descripción general de los conceptos centrales de los dos principales paradigmas de probabilidad para evaluar hipótesis: la NHST y la inferencia bayesiana. Mientras que la mayoría de los científicos utilizan ampliamente la primera, el planteamiento bayesiano se considera un moderno sistema de aprendizaje basado en datos que se aplica cada vez más en la arqueología (Buck y Mesón 2015: 205; Buck et al. 1996; Howson y Urbach 2006; Jaynes 2003; Otárola-Castillo y Torquato 2018; Otárola-Castillo et al. 2022).

PRUEBA DE SIGNIFICACIÓN DE LA HIPÓTESIS NULA

Como marco estadístico prevaeciente en la mayoría de las ciencias, la NHST permite a los profesionales utilizar sus datos para evaluar hipótesis. Este planteamiento tiene sus raíces en el desarrollo a principios del siglo XX de las pruebas de bondad de ajuste (Fisher 1922; Pearson 1900), el diseño experimental, los valores de p (Fisher 1925, 1935), los intervalos de confianza (IC) y las pruebas de hipótesis (Neyman y Pearson 1933: 294). Esta metodología se introdujo en la arqueología a mediados del siglo XX (por ejemplo, Binford 1964; Clarke 1968; Myers 1950; Spaulding 1953; Vesceilius 1960). Las aplicaciones de la NHST en arqueología continúan hoy, respaldadas por nuevos libros de texto estadísticos específicos de esa disciplina (por ejemplo, Banning 2020; Baxter 2003; Carlson 2017; Drennan 2010; Fletcher y Lock 2005; McCall 2018; Shennan 1997; Thomas 1986). Estas publicaciones brindan un tratamiento detallado de la NHST y sus procedimientos en el contexto de la arqueología (para una

introducción multidisciplinaria a la NHST, consulte, por ejemplo, Diez et al. 2019).

Sin embargo, en general, el paradigma de la NHST gira en torno al concepto de muestreo repetido teóricamente a largo plazo y el teorema central del límite (TCL; Diez et al. 2019: 172). El TCL informa el planteamiento de la NHST para la descripción y evaluación de hipótesis, e indica que dada una muestra lo suficientemente grande, en muchos casos, las estadísticas de resumen (por ejemplo, la media o desviación estándar) seguirán una distribución normal. Por ejemplo, después de muestrear la misma población varias veces, las medias de las muestras individuales se distribuirán normalmente. Esta distribución se conoce como muestreo o distribución "nula" de la estadística. Como este fenómeno ocurre frecuentemente, incluso si la variable original no tenía una distribución normal, este concepto se aplica a muchas situaciones y datos. El TCL vincula además las estadísticas de muestra con sus distribuciones nulas, como la media, a través de su error estándar. Según el TCL, el error estándar de la media de una muestra estima la desviación estándar de la distribución nula de la media. Esta cantidad puede calcularse dividiendo la desviación estándar de la muestra por el tamaño de la muestra.

El TCL es útil para los arqueólogos que a menudo toman muestras de una población objetivo: un grupo de individuos, artefactos, eventos, mediciones u otros fenómenos que desean estudiar. El objetivo es usar la muestra para probar hipótesis a priori sobre características cuantificables de la población muestreada. Los estadísticos se refieren a estas características como los parámetros de la población. Por ejemplo, los parámetros de media y desviación estándar de una población representan su tendencia central y variabilidad, respectivamente. Las estadísticas de muestra funcionan como estimaciones de los parámetros de la población y, por lo tanto, también se conocen como estimaciones de parámetros. Estas estadísticas se usan con la finalidad de evaluar hipótesis sobre los respectivos parámetros de población. La NHST requiere que los arqueólogos establezcan únicamente dos hipótesis: una hipótesis nula y una alternativa a evaluar. Las hipótesis nulas son afirmaciones cuantitativas de "ninguna diferencia" (diferencia = 0) entre un valor de parámetro hipotético y su estadística de muestra, o entre una estadística de muestra y su contraparte de otra muestra. Los arqueólogos suelen establecer tales hipótesis nulas para evaluar si una muestra estadística resultó de una población que tiene el valor del parámetro hipotético (es decir, una prueba de una muestra). Alternativamente, pueden desear saber si las estadísticas de dos muestras independientes se extrajeron de la misma población (es decir, una prueba de dos muestras).

Las hipótesis alternativas son declaraciones ordinariamente simples que niegan la hipótesis nula. Una vez que los arqueólogos establecen las hipótesis nula y alternativa, toman muestras de la población o "recolectan datos" y calculan las estadísticas de la muestra. Tenemos que señalar que el marco que ofrece la NHST procede asumiendo que la hipótesis nula es verdadera y posteriormente utiliza los datos de la muestra, resumidos por una estadística, para evaluar esa suposición. Para hacerlo, los arqueólogos utilizan la estadística de muestra para definir una estadística de prueba (con frecuencia, los valores z , t , ratios F - y los valores de χ^2 ; por ejemplo, Diez et al. 2019; Drennan 2009; Thomas 1986) y calculan la probabilidad de que un valor igual o más extremo que el estadístico de prueba pueda ocurrir bajo el supuesto de la hipótesis nula.

La probabilidad de la estadística de prueba, o valor p , a menudo se calcula con la ayuda de modelos de distribución de probabili-

dad, como la distribución normal. Estos modelos de probabilidad son conocidos también como funciones de verosimilitud. La verosimilitud es una función estadística que describe la probabilidad de la estadística de prueba que depende de los valores de los parámetros hipotéticos, por ejemplo, los asumidos por la hipótesis nula. Así como mostramos en el ejemplo ficticio a continuación, la función de verosimilitud normal es utilizada para calcular el valor p de una estadística de prueba de ratio z , suponiendo que la hipótesis nula es verdadera. Utilizando modelos de probabilidad similares, los arqueólogos aplican la NHST y calculan cantidades tales como valores p e intervalos de confianza (IC) para evaluar si la estadística de prueba rechaza la hipótesis nula.

Los IC están basados en el concepto de distribución nula del TCL. Los arqueólogos a menudo calculan los IC en dos contextos: (1) para realizar la NHST, calculan los IC de una estadística de prueba; y (2) para estimar la precisión de la estimación de un parámetro, calculan los IC de una estadística de muestra. En general, los IC de la estadística de prueba o de muestra se centran en su media, representan su distribución nula respectiva y se derivan usando el error estándar de su muestra. Recuerde que el error estándar de cualquier estadística es la desviación estándar de su distribución nula. Para la estadística muestral, esta distribución representa el rango de valores plausibles dentro de los cuales puede encontrarse el verdadero valor de los parámetros de la población.

Sin embargo, en el contexto de la estadística de prueba, el IC es el rango de valores posibles dentro del cual se encontrará la verdadera diferencia, asumida por la hipótesis nula. Dicho de otro modo, debido al TCL, aproximadamente 68 % de la distribución nula de la estadística de prueba capturará el verdadero valor de la diferencia, que la hipótesis nula supone que es 0. Asimismo, en el caso de un estadístico muestral, el 68 % de su distribución nula contendrá el verdadero valor del parámetro poblacional. Alternativamente, se puede desear una incertidumbre inferior al 68 % para la muestra o la estadística de prueba. En este caso, pueden calcularse rangos similares al error estándar que capturan el parámetro real o los valores de diferencia entre el 95 % y el 99 % de las veces, nuevamente, después de un muestreo teórico repetido. Estos rangos son los IC, y nos referimos a ellos en términos de su porcentaje: por ejemplo, como 95 % o 99 % IC. En el contexto de la NHST, los arqueólogos usan los IC de la estadística de prueba para rechazar o no una hipótesis nula. Si el valor de ninguna diferencia, 0, está dentro del IC de la estadística de prueba, entonces la hipótesis nula no se rechaza. Sin embargo, si 0 no está dentro del rango de IC de la estadística de prueba, los datos no respaldan la hipótesis nula y se rechaza a favor de la alternativa. Ofrecemos una última nota sobre la mecánica de los IC. Puede parecer tentador interpretar el IC del 95 % como una indicación de que el verdadero parámetro o diferencia de la población tiene una probabilidad de 0.95 de estar en el IC. Sin embargo, aunque algo confuso, la interpretación correcta del IC es que, con base en un muestreo repetido a largo plazo, el 95 % de los IC contendrá el verdadero parámetro o diferencia de la población.

Además de los IC, la NHST utiliza valores p como una señal empírica de la plausibilidad de la estadística de prueba, asumiendo que la hipótesis nula es verdadera. Los arqueólogos calculan los valores p basándose en la proporción de valores en la distribución nula igual y más extrema que el estadístico de prueba de la muestra. Por lo general, los valores estadísticos de prueba con un valor p menor o igual a una proporción de 0.05 (1 de 20 o 5 %) son considerados extremos. Es común que los arqueólogos evalúen rechazar la hipótesis nula utilizando un valor p de 0.05 como límite para el rechazo: cuanto más extremos son los datos,

menor es el valor p .

La comunidad científica en general se ha vuelto cada vez más crítica con la NHST (por ejemplo: Gelman 2006, 2018; Vidgen y Yasseri 2016). Los estadísticos y practicantes han señalado enérgicamente la arbitrariedad del umbral del valor p de 0.05 para la significación estadística (Cowgill 1977: 352; Vallengia y Fernández-Duque 2022; Wasserstein et al. 2019). Algunos argumentan que una formación estadística no adecuada puede llevar a los investigadores a malinterpretar los valores p (Hubbard 2011: 2624; McShane y Gal 2015). Una consecuencia de no comprender completamente el concepto de valores p , por ejemplo, es que algunos investigadores confunden el significado práctico, o relevancia, con el significado estadístico. En particular, es posible que los efectos que son prácticamente insignificantes, irrelevantes o poco interesantes den como resultado valores p pequeños (por ejemplo, Aarts et al. 2012; Johnson 1999; Kramer et al. 2016; McCall 2018: 90-93; Wolverton et al. 2016). En un caso, mientras investigaban los efectos de la competencia entre hermanos en los patrones de crecimiento de los niños mayas, Kramer y colegas (Kramer et al. 2016) encontraron que los efectos del tamaño de la familia en el crecimiento de los niños eran estadísticamente significativos, pero “de poca importancia para la salud o el estado físico de la primera infancia”. Aquí, interpretar el límite del valor p de 0.05 como demográficamente importante habría llevado a conclusiones incorrectas.

En otros casos, los investigadores han confundido los valores p con la tasa de error de tipo I, α . El valor p es la probabilidad de que la estadística de prueba pueda ocurrir bajo la hipótesis nula; α es la probabilidad de rechazar la hipótesis nula cuando es verdadera (Hubbard 2011). Históricamente, estas dos cantidades estadísticas pertenecen a filosofías de la NHST en competencia (Fisher 1925; Neyman y Pearson 1933). Neyman y Pearson desarrollaron el concepto de error tipo I en el contexto del diseño de experimentos infinitamente repetibles, donde α define la probabilidad de que un análisis no encuentre una diferencia entre dos hipótesis cuando existe una genuina. Contrariamente, el valor p de Fisher, estima empíricamente si un conjunto específico de observaciones se ajusta a una hipótesis nula específica. Estas dos cantidades tienen fundamentos teóricos y relaciones completamente distintas con las observaciones reales. Por ejemplo, α no está relacionado con las observaciones y el valor p no está influenciado por las hipótesis alternativas bajo consideración. Desafortunadamente, la práctica típica de la NHST puede llevar a los investigadores a asociar directamente los dos conceptos, complicando los esfuerzos para proporcionar definiciones e interpretaciones razonables (Hubbard y Bayarri 2003). El mal uso de los valores de p y la significación estadística, debido a un malentendido (por ejemplo, Thiess et al. 2015) o intención (Chuard et al. 2019; Head et al. 2015), puede dar lugar a la llamada crisis de replicación científica (Ioannidis 2005) que empieza a alcanzar a la ciencia arqueológica (Bayliss y Marshall 2019; Marwick 2017; McPherron et al. 2021).

Incluso teniendo en cuenta estos matices, la interpretación de los conceptos de la NHST, como los valores p , la significación estadística, las pruebas de hipótesis y los IC, no es sencilla. Las declaraciones sobre las estadísticas de la muestra (errores estándar e IC) están basados en un muestreo repetido hipotético, que es difícil de concebir en situaciones no experimentales o, como en arqueología, donde la replicación real es difícil o incluso imposible de lograr. En términos de evaluación, aunque la mayoría de los investigadores en general pueden entender cómo interpretar un valor p significativo en el contexto de rechazar una hipótesis nula, el significado de un valor p no significativo puede causar

confusión. El desconcierto podría verse exacerbado porque no existe un mecanismo para “aceptar” o “verificar” una hipótesis nula. Este malentendido crítico de la NHST puede llevar a algunos a interpretar un valor p no significativo como una aceptación de su hipótesis nula en vez de rechazarla (Greenland et al. 2016). Sin embargo, la producción de conocimiento en el paradigma de la NHST se centra en rechazar las hipótesis nulas, en vez de aceptar las hipótesis nulas o alternativas. Para ser justos, el lenguaje de la NHST es confuso. Por ejemplo, afirmar que una hipótesis nula no pudo ser rechazada es un triple negativo, lo que significa que “la hipótesis de no diferencia no fue no aceptada”. Tal lenguaje intrincado incluido en la NHST ofusca la relación entre las hipótesis de valor p nula y alternativa.

Además, el papel de la hipótesis alternativa y su conexión con el valor p tampoco están claros y a menudo se interpretan incorrectamente (Benjamin y Berger 2019; Cohen 1994). Como resultado, la inferencia utilizando estadísticas de NHST tradicionales puede ser difícil, especialmente cuando un estudio desea discernir entre múltiples hipótesis de trabajo (por ejemplo, Chamberlin 1965; Gelman et al. 2012), tal como cuando dos o más hipótesis no logran ser rechazadas. En teoría, tales hipótesis son consistentes con los datos. Sin embargo, clasificar múltiples hipótesis nulas no rechazadas es difícil, si no imposible. Una forma de clasificarlos puede ser utilizar los valores p de las hipótesis. Después de todo, el valor p es una métrica continua que media el rechazo de hipótesis y la falta de rechazo. Sin embargo, los estadísticos desaconsejan este procedimiento (Hubbard y Lindsay 2008; McShane et al. 2019) porque la magnitud del valor p no refleja el peso de la evidencia de una hipótesis sobre otra. En consecuencia, la NHST tradicional no ofrece un procedimiento sencillo para comparar más hipótesis nulas “no rechazadas”.

ESTADÍSTICAS BAYESIANAS

La inferencia bayesiana ofrece un planteamiento alternativo con varias ventajas sobre la NHST. En primer lugar, las estadísticas bayesianas permiten a los científicos utilizar datos para asignar probabilidades a las estimaciones de sus parámetros e hipótesis, lo que facilita una comparación más directa de las hipótesis contrapuestas. En segundo lugar, mientras que la NHST utiliza solamente datos nuevos para hacer inferencias, un marco bayesiano permite combinar tanto datos nuevos como información existente. Como detallamos a continuación, esta característica se parece más a los procesos de toma de decisiones de los científicos y es probablemente una de las razones clave por las que estos, incluidos los antropólogos y arqueólogos, utilizan cada vez más la inferencia bayesiana para evaluar sus hipótesis.

El teorema de Bayes deriva su nombre del reverendo Thomas Bayes (1763), un ministro presbiteriano inglés y matemático que investigó problemas de probabilidad que involucraban probabilidades condicionales y previas (definidas a continuación). Sin embargo, no fue hasta finales de 1900 que el planteamiento bayesiano de la inferencia estadística se popularizó en la ciencia (Bellhouse 2004). Aunque los arqueólogos comenzaron notoriamente a adoptar estadísticas bayesianas para evaluar hipótesis en la década de 1990 (por ejemplo, Buck et al. 1996; Cowgill 1993), pueden encontrarse aplicaciones anteriores dispersas en la literatura de esa disciplina a partir de la década de 1970 (Doran et al. 1975; Fisher 1987; Freeman 1976; Salmon 1982: 51-55; Thomas 1986). Hoy en día, los científicos, incluidos los antropólogos y arqueólogos que encuentran ventajoso este planteamiento, están aplicando cada vez más las estadísticas bayesianas para evaluar sus hipótesis con datos (Gelman et al. 2020;

McElreath 2020; Naylor y Smith 1988; Otárola-Castillo y Torquato 2018; Otárola-Castillo et al. 2022).

Una ventaja de la inferencia bayesiana es que permite incorporar información previa o experta sobre hipótesis en los análisis estadísticos. Como mostramos en nuestro ejemplo a continuación, el conocimiento previo de un arqueólogo o un conjunto de arqueólogos y otros expertos puede ser muy valioso ya que “dependemos mucho de la información previa para ayudarnos a evaluar el grado de plausibilidad en un nuevo problema” (Jaynes 2003: 6). Incluir formalmente experiencia previa o información de expertos en análisis estadísticos para “actualizar” nuestro estado de conocimiento es un proceso de aprendizaje natural y mejora las inferencias hechas mediante la aplicación de la NHST (Cowgill 2001). Para lograr esto, los practicantes de la inferencia bayesiana convierten el conocimiento previo en probabilidades previas y las utilizan junto con sus distribuciones como parte de los análisis estadísticos. Una vez que los analistas determinan sus distribuciones de probabilidad previas, como con la NHST, pueden observar nuevos datos para probar su (o sus) hipótesis. En este contexto, la verosimilitud de estos se combina con (o se pondera por) la previa para obtener la probabilidad posterior bayesiana. Esta última es la probabilidad de la hipótesis dada la verosimilitud de los datos observados y el conocimiento previo (Buck et al. 1996). Como comentamos más detalladamente a continuación, el proceso bayesiano es particularmente útil en situaciones en las que solamente se obtienen pequeñas cantidades de datos, como suele ser el caso en arqueología.

En casos simples, determinar la probabilidad posterior y su distribución es relativamente sencillo. Sin embargo, el cálculo subyacente a casos más complejos es imposible de resolver sin la aplicación de nuevos métodos de simulación. En particular, los algoritmos de Cadena Markov Monte Carlo (MCMC por sus siglas en inglés) han facilitado el progreso en los análisis bayesianos. La simulación MCMC es una combinación de muestreo de Monte Carlo y cadenas de Markov. El muestreo de Monte Carlo se utiliza para estimar cantidades difíciles de calcular a partir de la distribución desconocida de una variable aleatoria observada. Las cadenas de Markov son una serie estocástica de eventos asociados entre sí, donde la probabilidad de un nuevo evento depende únicamente del estado del último. Juntas, estas características del muestreo de Monte Carlo y las cadenas de Markov son esenciales para encontrar la distribución de probabilidad posterior de problemas complejos. Hoy en día, las variaciones del algoritmo MCMC original (Metropolis et al. 1953), como Metropolis-Hastings, Gibbs, hamiltoniano y otros métodos, se utilizan ampliamente, lo que facilita una amplia aplicación del paradigma bayesiano (por ejemplo, Dunson y Johndrow 2020; Gilks et al. 1995; Howson y Urbach 2006: xi; Robert y Casella 2011).

Para contextualizar todavía más la aplicación de las estadísticas bayesianas, proporcionamos un ejemplo ficticio que ilustra cómo puede utilizarse este marco probabilístico para resolver un problema de investigación arqueológica idealizado. Para hacer esto, elegimos usar un caso simulado¹ en vez de un estudio de caso real para evitar las complejidades de los procesos de formación de sitios y el sesgo de muestreo. Este ejemplo inventado también ayuda a centrar la atención en aspectos específicos de la inferencia bayesiana que consideramos muy instructivos. Este caso de la “cultura Monico y el arqueólogo bayesiano” demuestra cómo se pueden hacer inferencias utilizando datos e información previa sobre una hipótesis, cómo evaluar la incertidumbre que rodea a esta, por qué este plantamiento parece menos ambiguo que la NHST y, por lo tanto, por qué se está volviendo cada vez más popular.

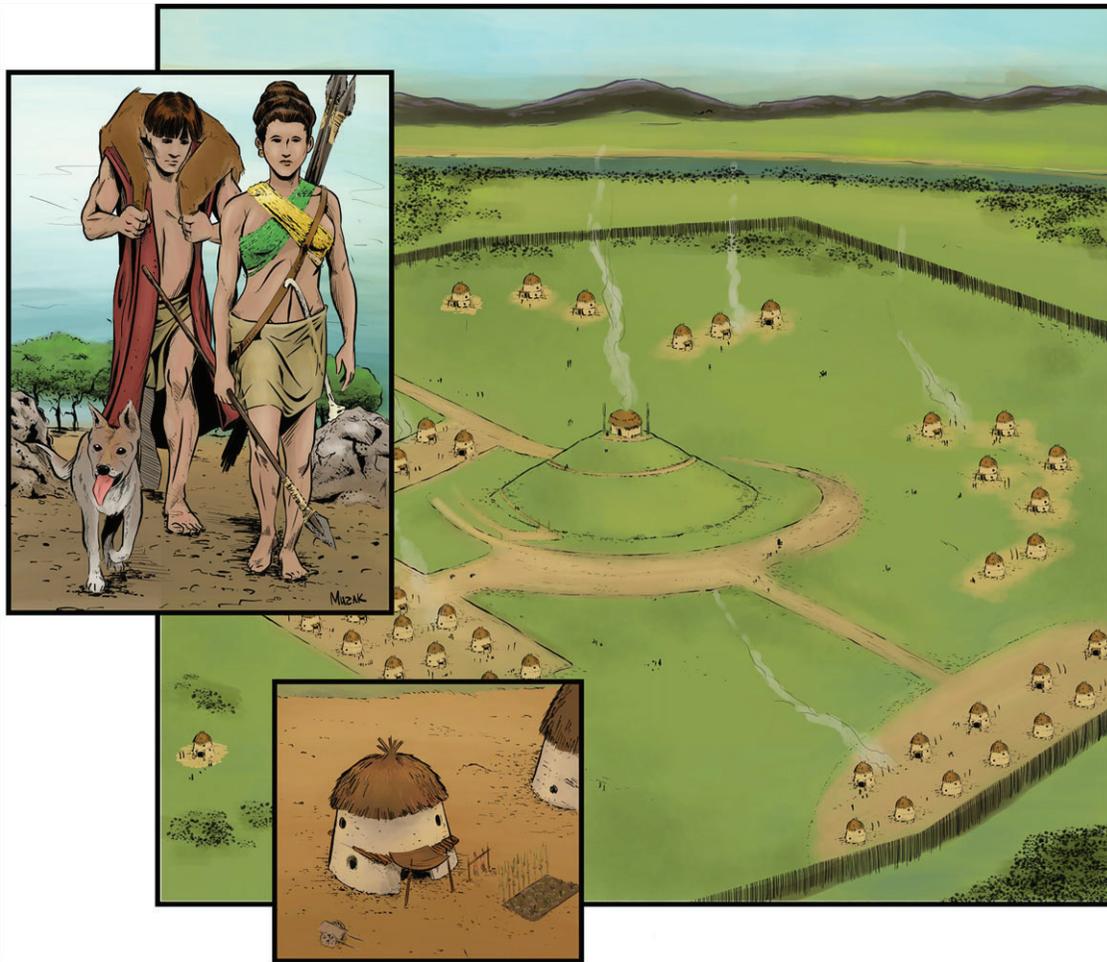


FIGURA 1. Reconstrucción del sitio Monico-1 perteneciente a la cultura arqueológica ficticia de Monico.

LA CULTURA MONICO: UNA APLICACIÓN SIMPLIFICADA DE LAS ESTADÍSTICAS BAYESIANAS

La “cultura Monico” hace referencia a un grupo ficticio de personas que podrían haber vivido entre el presente etnográfico y hace mucho tiempo en múltiples entornos ambientales y contextos socioculturales en todo el mundo. El registro arqueológico imaginario de Monico es bien conocido. En general, su cultura material refleja patrones de economías de recolección, agricultura y pastoreo. Sus dinámicas socioculturales son amplias: van desde prácticas igualitarias exhibidas en campamentos altamente móviles hasta una mayor complejidad social derivada de asentamientos más permanentes. Algunos expertos en esta cultura argumentan que los asentamientos más tardíos muestran evidencia de producción intensiva de alimentos, comercio de productos exóticos y una organización política altamente centralizada y administrada por una élite cada vez más jerárquica (Figura 1).

Un famoso arqueólogo bayesiano —una autoridad en Monico— ha excavado un sitio del período posterior al contacto asociado con esta cultura. El trabajo de excavación en Monico-1 produjo un conjunto de fauna impresionante entre la cultura material ampliamente diversa. La arqueofauna está compuesta por dos

especies de animales: perro y coyote, representados por esqueletos completos. En consecuencia, en este informe, el arqueólogo utiliza el término “individuo” para referirse a perros o coyotes completos. Asimismo, cuando el arqueólogo menciona “el número de” perros o coyotes, se refiere a un conteo de individuos completos de la respectiva especie. Hasta el momento, el arqueólogo ha identificado 100 de estos individuos y los ha asignado a sus respectivas especies. Con base en las observaciones, el conjunto está compuesto por 71 perros y 29 coyotes (Figura 2).

Sin embargo, el arqueólogo también ha excavado un fragmento de hueso que es difícil de identificar y quiere saber cuál es la especie más probable a la que pertenecería este fragmento.

El arqueólogo define “probabilidad” como la frecuencia relativa o proporción de veces que ocurre un evento. Sobre la base únicamente de los datos, la probabilidad, P , de que el perro permanezca en el conjunto es:

$$P(\text{Perro}) = \frac{71}{100} = 0.71,$$

mientras que la probabilidad de que quede un coyote es:

$$P(\text{Coyote}) = \frac{29}{100} = 0.29.$$

Dadas estas probabilidades, es razonable que el arqueólogo crea que es más probable que el espécimen óseo no identificable sea



FIGURA 2. El arqueólogo bayesiano y su equipo excavan el sitio Monico-1.

de un perro. Sin embargo, se muestra escéptico. Además, como estudioso de Monico, ha recopilado información etnográfica sobre el comportamiento de su población, particularmente sobre sus tabúes alimenticios. Los relatos históricos revelan que los monico alguna vez mantuvieron perros en sus aldeas para cazar coyotes. Debido a que la base de la subsistencia tradicional de Monico dependía de la caza del coyote, los perros desarrollaron relaciones especiales con sus dueños. En consecuencia, los monico llegaron a tratar a sus perros con respeto, como lo harían con otras personas.

La tradición oral transmitida de generación en generación ha documentado que se consideraba a los perros hermanos cercanos de las personas. En particular, se sabe que la cultura Monico tenía como tabúes matarlos o comerlos. Sin embargo, las narraciones orales también han revelado que los monico comían perros en tiempos de grave escasez de alimentos. Con esta información adicional o “previa”, el arqueólogo decide observar los esqueletos más de cerca para comprobar la presencia de marcas de matanza (es decir, huellas de corte) en los restos del perro. Luego tabula esta información adicional sobre los huesos recuperados bajo dos condiciones: 1) las marcas de matanza están presentes y 2) las marcas de matanza están ausentes. La [Tabla 1](#) muestra las frecuencias de marcas de matanza en los esqueletos de cada especie.

Para convertir estos datos en una tabla de probabilidad, el arqueólogo estandariza (o divide) todos los valores por el número total de observaciones (100 en este caso). Las celdas internas (fuente oscura, sombreado claro) en la [Tabla 2](#) proporcionan las probabilidades de que las marcas de matanza y las

especies ocurran juntas o en conjunto, lo que se conoce como *probabilidades conjuntas*.

Los valores en los márgenes derecho e inferior de la [Tabla 2](#) se denominan adecuadamente *probabilidades marginales*. Estos representan la presencia y ausencia de marcas de matanza (a la derecha) y la especie identificada (abajo). Los totales marginales son las probabilidades totales de cada espacio subdividido (especie o marca de matanza). Por definición, todas las probabilidades se encuentran en el rango de 0 a 1, y la suma total de las filas o columnas marginales (es decir, la suma de todos los resultados marginales) debe ser 1.

En este punto, el arqueólogo se enfoca en el espécimen de hueso no identificable y encuentra en él varias marcas de matanza. Puede utilizar esta información adicional para obtener una ventaja inferencial al tener en cuenta o condicionar la presencia de marcas de matanza, un proceso llamado *condicionamiento*. Así condiciona la especie identificada a la presencia o ausencia de marcas de matanza. Este procedimiento también se conoce como subconjunto o estratificación de la variable “especies identificadas” por la presencia o ausencia de las huellas mencionadas.

Naturalmente, el arqueólogo pregunta: “¿Cuál es la probabilidad de que el espécimen óseo no identificable sea de un perro en comparación con la probabilidad de que sea de un coyote, dado que hay marcas de matanza en los huesos del individuo?”. Él observó 32 esqueletos de animales de Monico-1 con marcas de matanza presentes: en 9 perros y 23 coyotes. Con esos datos puede calcular las probabilidades de que el individuo pertenezca a una especie u otra, dado que hay marcas de matanza (los

TABLA 1. Frecuencias de individuos y marcas de matanza observadas en Monico-1

		Individuos de Cada Especie		Total de marcas de matanza
		Perro	Coyote	
Marcas de matanza	Presente	9	23	32
	Ausente	62	6	68
Total de Individuos		71	29	Total = 100

Nota: Tenga en cuenta que si bien la mayoría de las marcas de matanza corresponden a huesos de coyote, 9 de los 71 huesos de perro también muestran signos de matanza.

estadísticos usan el símbolo “|” para referirse a “dado que” y para indicar que se está produciendo un condicionamiento). Para un perro, la probabilidad es:

$$P(\text{Perro} \mid \text{Marca de matanza presente}) = \frac{9}{32} = 0.28,$$

mientras que la probabilidad de que un individuo con marcas de matanza pertenezca a la especie coyote es:

$$P(\text{Coyote} \mid \text{Marca de matanza presente}) = \frac{23}{32} = 0.72.$$

Por lo tanto, después de observar las marcas de matanza en el hueso individual (no identificado), el arqueólogo puede afirmar que la probabilidad de que provenga de un coyote es de 0.72. En otras palabras, tiene un 72 % de certeza de que el hueso pertenece a un coyote.

Unos días después, un reportero de un periódico local se enteró de una excavación arqueológica en curso en otro sitio cercano de la aldea de Monico, llamado Monico-2. Sus fuentes le revelan que allí también se han recuperado restos de fauna. Debido a que el arqueólogo es un conocido experto en los hábitos alimenticios de Monico, el reportero se pone en contacto con él y le comunica que el conjunto de fauna de Monico-2 está compuesto exclusivamente por restos de especies de perros.

Aunque los investigadores de Monico-2 todavía no han realizado un análisis completo de la fauna, el reportero pregunta al arqueólogo qué tan probable es que los monico estuvieran matando y comiendo perros en el nuevo sitio. Por ahora, el arqueólogo ha estimado las probabilidades de encontrar marcas de matanzas asociadas con cada especie animal en base a la experiencia en Monico-1. Para hacer una inferencia probabilística sobre el comportamiento en el nuevo sitio, condiciona las “especies identificadas” en lugar de la “presencia de marcas de matanza”. De los 71 perros identificados en Monico-1, el arqueólogo observó 9 con marcas de matanza y 62 sin ellas. Esto significa que, según la evidencia de Monico-1, la probabilidad de encontrar evidencia de matanza de perros es:

$$P(\text{Marca de matanza presente} \mid \text{Perro}) = \frac{9}{71} = 0.13,$$

mientras que la probabilidad de que no haya evidencia de matanza en perros es:

$$P(\text{Marca de matanza ausente} \mid \text{Perro}) = \frac{62}{71} = 0.87.$$

Después de pensar un momento, el arqueólogo dice al reportero que, basado en el conocimiento de Monico-1, la probabilidad

TABLA 2. Probabilidades conjuntas de individuos y marcas de matanza observadas

		Individuos de cada especie		Marcas marginales de matanza
		Perro	Coyote	
Marcas de matanza	P(Presente)	0.09	0.23	0.32
	P(Ausente)	0.62	0.06	0.68
Especies marginales		0.71	0.29	Total = 1

Nota: Estas cantidades describen la probabilidad de identificar una especie y observar marcas de matanza en los huesos de la misma; por ejemplo, $P(\text{Coyote y marcas de matanza presentes})$ es 0.23, o 23 %.

de que los huesos de perro de Monico-2 hayan resultado de actividades dietéticas es relativamente baja, alrededor del 13 %. Este cálculo se basa en el teorema de Bayes, así como en la información sobre la relación de Monico con sus perros y las prácticas de matanza en Monico-1.

Qué es el teorema de Bayes?

El teorema de Bayes es la formalización algebraica del trabajo de tabla probabilística que realizamos en la sección anterior utilizando un evento discreto. El teorema es más útil cuando se conoce un enunciado de probabilidad condicional y se desea obtener su enunciado condicional inverso. Por ejemplo, del modelo anterior, sabemos que $P(\text{Marca de matanza presente} \mid \text{Perro}) = 0.13$. Si deseamos conocer el enunciado condicional inverso $P(\text{Perro} \mid \text{Marca de matanza presente})$, podemos calcularlo usando:

$$P(\text{Perro} \mid \text{Marca de matanza presente}) = \frac{P(\text{Marca de matanza presente} \mid \text{Perro}) \times P(\text{Perro})}{P(\text{Marca de matanza presente})}$$

Las tablas 1 y 2 proporcionan los valores necesarios para sustituir esta expresión de manera que:

$$P(\text{Perro} \mid \text{Marca de matanza presente}) = \frac{\left(\frac{0.09}{0.71}\right) \times 0.71}{0.32} = 0.28.$$

Cuando se generaliza, el algoritmo aplicado aquí es conocido como teorema de Bayes. Por lo general, se ejemplifica considerando dos eventos relacionados: A y B. En pocas palabras, el teorema de Bayes establece que:

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

En este caso, para obtener la probabilidad condicional de A dado B, $P(A|B)$, hay que dividir la probabilidad conjunta de A y B, $P(A \text{ y } B)$, por la probabilidad marginal de B, $P(B)$. El producto de $P(B|A)$ y $P(A)$ es la probabilidad conjunta, $P(A \text{ y } B)$. La fórmula entonces se generaliza a:

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

TABLA 3. Frecuencias de individuos y marcas de matanza observadas en Monico-2

		Individuos de cada especie		Total de marcas de matanza
		Perro	Coyote	
Marcas de matanza	Presente	9	1	10
	Ausente	1	0	1
Total de Individuos		10	1	Total = 11

Nota: Tenga en cuenta el pequeño número total de individuos y la muestra particularmente pequeña de coyotes.

donde la probabilidad conjunta se divide por el $P(B)$ marginal. Los estadísticos llaman a $P(A|B)$ la probabilidad posterior de A dado B ; $P(B|A)$ el condicional inverso (o verosimilitud) de B dado A ; y $P(A)$ la probabilidad previa de A .

EL ARQUEÓLOGO BAYESIANO CONTINUADO

Unos días después, el reportero adquiere más información de las excavaciones que continúan realizándose en el pueblo de Monico-2. Las frecuencias y probabilidades conjuntas se describen en las Tablas 3 y 4 a continuación. El reportero informa muy emocionado al arqueólogo que sus colegas habían recuperado 10 perros y que todos menos uno tenían marcas de matanza. Por el contrario, los arqueólogos del sitio Monico-2 habían recuperado *solamente* un coyote que exhibía marcas de matanza en los restos. Los investigadores de Monico-2 utilizaron una estadística de NHST apropiada, la prueba z unilateral para proporciones (Diez et al. 2019: 194-197), con corrección de continuidad, para probar si la tasa de matanza de perros observada (9/10) era estadísticamente significativamente superior al 50 %, la hipótesis nula predeterminada en esta prueba. Los arqueólogos de Monico-2 rechazaron la hipótesis nula con un valor de $p < 0.05$ (ratio $z = 2.21$, mean = 5, sdev = 1.58, $p = 0.013$). Debido al pequeño tamaño de la muestra, también realizaron una prueba binomial unilateral, que arrojó resultados en línea con aquellos de la prueba z (éxitos = 9, intentos = 10, $p = 0.01074$). Con base en estos resultados estadísticamente significativos, los arqueólogos de Monico-2 dijeron al reportero que la mayoría de los perros fueron sacrificados en el sitio. Además, según este, los arqueólogos también sugirieron que la evidencia y los resultados del análisis estadístico indicaron que la gente de la aldea Monico-2 incluía perros como parte importante de su dieta. A la luz de esta evidencia, el reportero comienza a cuestionar el registro etnográfico sobre los tabúes dietéticos de los monico.

El arqueólogo de Monico-1 echa un vistazo rápido a las tablas, hace algunos cálculos y sostiene que la probabilidad de que los aldeanos de Monico-2 hayan sacrificado a sus perros ahora es incluso menor, especialmente en comparación con la nueva probabilidad de matanza de coyotes, que es ligeramente superior. Sin embargo, insiste en esperar una muestra más grande antes de emitir conclusiones firmes. Incrédulo, el reportero pide una explicación de por qué el arqueólogo cuestiona las significativas pruebas de hipótesis nulas realizadas por los arqueólogos de Monico-2. El arqueólogo

TABLA 4. Probabilidades conjuntas de individuos y marcas de matanza observadas en Monico-2

		Individuos de cada especie		Marcas marginales de matanza
		Perro	Coyote	
Marcas de matanza	P(Presente)	0.82	0.09	0.91
	P(Ausente)	0.09	0	0.09
Especies marginales		0.91	0.09	Total = 1

Nota: Tenga en cuenta la mayor proporción de huesos de perro con marcas de matanza en comparación con la muestra de Monico-1.

mira al reportero y dice: "Bueno, los procedimientos de la NHST como la prueba z únicamente consideran datos nuevos. Desafortunadamente, estos métodos no tienen en cuenta toda la información disponible, nueva y previa, sobre la subsistencia de Monico. Personalmente —prosigue el arqueólogo— intento no formar mis opiniones basándome únicamente en nuevos datos. Más bien, utilizo nuevos datos con el fin de actualizar mis opiniones hechas con conocimiento previo, por ejemplo, del sitio Monico-1". Luego, el arqueólogo guía al reportero a través de las tablas y comienza a explicar cómo hace su inferencia usando el teorema de Bayes.

El arqueólogo explica que las probabilidades posteriores de matanza de perros y coyotes extraídas del conjunto de fauna (mucho más grande) de Monico-1 se han convertido en información "previa" sobre las probabilidades de que los aldeanos de Monico mataran perros y coyotes. Estas cantidades pueden ser representadas por:

$$P(\text{Marca de matanza presente} \mid \text{Perro})_{\text{Monico-1}} = \frac{9}{71} = 0.13,$$

y

$$P(\text{Marca de matanza presente} \mid \text{Coyote})_{\text{Monico-1}} = \frac{23}{29} = 0.79.$$

El conocimiento del arqueólogo sobre el grado en que los aldeanos de Monico-1 sacrificaron perros y coyotes puede actualizarse en una nueva iteración del teorema de Bayes que incluye los datos de Monico-2. Para dar cuenta del contexto arqueológico del que derivan los cálculos, el arqueólogo añade los subíndices *Monico-1* y *Monico-2* a los términos de la ecuación, de la siguiente manera:

$$\begin{aligned} &P(\text{Marca de matanza presente} \mid \text{Perro})_{\text{Monico-2}} \\ &= \frac{P(\text{Perro} \mid \text{Matanza})_{\text{Monico-2}} \times P(\text{Perro} \mid \text{Matanza})}{P(\text{Perro})_{\text{Monico-2}}} \\ &= \frac{0.82}{0.91} \times 0.13 \\ &= \frac{0.1066}{0.91} = 0.13. \end{aligned}$$

Añadir los datos de perros de Monico-2 hace que la probabilidad de matarlos disminuya ligeramente (de 0.127 a 0.126, pero redondeada a 0.13). Puede utilizarse la misma operación empleando los datos anteriores de la primera excavación y los

nuevos coyotes:

$$\begin{aligned}
 &P(\text{Marca de matanza presente} \mid \text{Coyote})_{\text{Monico-2}} \\
 &= \frac{P(\text{Coyote} \mid \text{Matanza})_{\text{Monico-2}} \times P(\text{Coyote} \mid \text{Matanza})}{P(\text{Coyote})_{\text{Monico-2}}} \\
 &= \frac{\frac{0.09}{0.91} \times 0.79}{0.09} = 0.87.
 \end{aligned}$$

En este caso, después de actualizar los datos, la nueva probabilidad posterior de matanza de coyotes también es mayor (cambiando aún más respecto a la probabilidad previa que en el caso de los perros). El arqueólogo explica esto al reportero e insta a la cautela dado que los datos y las probabilidades resultantes del sitio original se derivaron de una muestra de 100 individuos, mientras que la selección actual representa un total de solamente 11. Aunque los cálculos de probabilidad son correctos, sería prudente esperar más datos, ya que la excavación de Monico-2 está en curso. Sin embargo, el análisis bayesiano del arqueólogo sugiere que, en este punto, no deberíamos esperar marcas de matanza en ningún perro recién descubierto en ese sitio.

VINCULACIÓN DEL TEOREMA DE BAYES A DATOS E HIPÓTESIS

El estudio de caso de Monico proporciona un ejemplo tangible de los diferentes componentes de un análisis bayesiano, incluida la estimación de la probabilidad de un evento y la probabilidad de un evento dado otro (usando los datos actualmente disponibles), junto con los conceptos clave de verosimilitud, probabilidades previas y posteriores, y cómo actualizar nuestro conocimiento utilizando el anterior posterior bayesiano como el nuevo previo. Aunque el procedimiento ejemplificado aquí es específico para los datos de conteo arqueológico, el teorema de Bayes es muy general y puede ser útil para una amplia variedad de datos y procesos de generación de los mismos. Esta sección generaliza el teorema de Bayes a una variedad de otros escenarios científicos.

Anteriormente mencionamos que los científicos bayesianos utilizan los datos disponibles (D) para asignar probabilidades a declaraciones o hipótesis (H) sobre una población. El enunciado $P(H|D)$, es decir, la probabilidad de la hipótesis dados los datos formaliza esta relación. En nuestro ejemplo de los sitios de Monico, el arqueólogo estaba tratando de calcular la probabilidad de que su gente matara perros y coyotes (las hipótesis) dada la cantidad de marcas de corte en sus huesos (los datos disponibles). Para operacionalizar esta declaración en el contexto de datos e hipótesis, el teorema de Bayes funciona de la siguiente manera:

$$P(H|D) = \frac{P(D|H) \times P(H)}{P(D)},$$

donde $P(H|D)$ es la probabilidad posterior de la hipótesis dados los datos; $P(D|H)$ es la probabilidad de los datos dada la hipótesis (o la verosimilitud) de los datos observados; $P(H)$ es la probabilidad previa de la hipótesis (antes de que se recopilaran los datos); y $P(D)$ es la probabilidad de los datos disponibles (de todos los valores posibles de los datos). Alternativamente, generalizando y utilizando un vernáculo estadístico más moderno, esta operación puede expresarse como:

$$\text{Posterior} = \frac{(\text{Verosimilitud} \times \text{Previa})}{P(\text{Datos})}$$

De esta manera, la estadística bayesiana ofrece un marco estadístico alternativo para actualizar y evaluar hipótesis a través de un mecanismo que obtiene información a posteriori sobre el posterior de interés a partir de los datos, un modelo estadístico (expresado como verosimilitud) e información previa adecuadamente formulada. En otras palabras, con una declaración explícita de nuestra información previa, un modelo estadístico claramente definido y el deseo de actualizar nuestra comprensión, el teorema de Bayes nos brinda un marco probabilístico para hacer interpretaciones.

Además de la naturaleza coherente y explícita del marco, hay otra característica atractiva del paradigma bayesiano, a saber, que nos permite aprender de la experiencia. Los antecedentes permiten la contextualización explícita de conocimientos o creencias previas sobre el tema que se investiga (Buck et al. 1996; Cowgill 1993). El uso de conocimientos previos debería ser una tendencia natural para los arqueólogos. Como Buck y colegas (1996), ellos aplican conocimientos previos a menudo, por ejemplo, al inferir la función de los artefactos recién descubiertos mediante su asociación con otros hallados previamente y sus características. De manera similar, el arqueólogo de nuestro ejemplo pudo contextualizar los datos del sitio Monico-2 basándose en las observaciones de Monico-1. Pocos marcos interpretativos ofrecen una estructura clara para actualizar las creencias a la luz de la nueva información y, sin embargo, esta es una parte muy importante de la mayoría de los enfoques intuitivos para aprender sobre el mundo en el que vivimos. Además, la información posterior de hoy (basada en datos actuales e información previa) se encuentra en una forma adecuada para convertirse en información previa para trabajos futuros siempre y cuando haya más datos disponibles.

Desde inferencias sobre puntos discretos hasta distribuciones de datos

Hasta ahora, el ejemplo ha mostrado cómo se puede aplicar la inferencia bayesiana a hipótesis definidas por declaraciones sobre eventos discretos. En el caso ficticio presentado, las hipótesis estaban representadas por afirmaciones sobre si los restos de fauna observados eran el resultado de una matanza. Los datos asignaron probabilidades a cada hipótesis, indicando así la cantidad o grado de creencia en ellas. Estos datos fueron eventos discretos de solamente dos sitios. Sin embargo, aunque la proporción de huesos de perro sacrificados son los resultados del mismo proceso de comportamiento (matanza), es probable que estos valores varíen de un sitio a otro.

En consecuencia, muchos arqueólogos podrían desear comparar sus datos de un solo sitio con el universo de sitios conocidos. En este caso, las hipótesis a evaluar se caracterizan por los valores de los parámetros de un modelo de probabilidad. Si bien lo mencionamos anteriormente, en este punto vale la pena recordar que tales parámetros describen ciertas características de una muestra o población. Para los arqueólogos, los parámetros más comunes son los que miden la tendencia central, como la media o la mediana. La inferencia bayesiana se puede aplicar utilizando otros parámetros, así como la distribución completa de la información posterior, previa y de datos. Estos suelen estar representados por modelos de probabilidad. Es muy posible que el más conocido sea el de probabilidad normal, en el que la distribución de probabilidad tiene una forma de campana

simétrica alrededor de un único valor medio. Cuando se trata de datos (de muestra) y modelos asociados de probabilidad, es convencional utilizar el símbolo romano x para representar los datos observados (o de muestra) y el símbolo griego θ (*theta*) para representar el parámetro (o conjunto múltiple de parámetros) del modelo de población que estamos tratando de conocer. Dado x y un modelo con parámetro(s) θ , podemos volver a formular el teorema de Bayes y sus tres componentes (la *verosimilitud*, la *previa* y la *posterior*) en el contexto de las distribuciones de datos y sus modelos de probabilidad.

La *verosimilitud* es una función estadística, o una expresión matemática, que asocia cantidades de datos individuales con sus respectivos valores de probabilidad. Su forma está determinada por el modelo de probabilidad específico que se utiliza, pero, en términos generales, está representada por $P(x|\theta)$, es decir, la distribución de probabilidad de los datos recién observados condicionados por los parámetros. En consecuencia, la *verosimilitud* es la probabilidad de observar valores de datos particulares dados algunos valores específicos (o hipotéticos) de los parámetros desconocidos. Por lo tanto, esta es una declaración formal de la relación entre los parámetros sobre los que queremos aprender y los datos que recopilamos.

La *previa* también es una función y puede representarse por $P(\theta)$. Es una declaración de lo que sabemos sobre la distribución de probabilidad de los parámetros antes de que se recolecten nuevos datos. En términos simples, podemos pensar en esto como la probabilidad que asignamos a la observación de valores específicos de los parámetros desconocidos en función de lo que sabíamos antes de observar los datos. Esta es una declaración formal de nuestro conocimiento antes de recopilar los datos más recientes.

Lo *posterior* es lo que queremos obtener: una combinación de la información contenida en los nuevos datos, la *verosimilitud* y la *previa*. El posterior está representado por $P(\theta|x)$. Como se presentó en la sección anterior, esta es la probabilidad de la hipótesis dados los datos, o $P(H|D)$. Es la distribución de probabilidad de los parámetros del modelo condicionados a los datos. En términos simples, podemos pensar en esto como la probabilidad que asignamos a valores específicos o hipotéticos de los parámetros desconocidos después de observar nuevos datos. En este contexto, podemos expresar el teorema de Bayes como:

$$P(\theta|x) = \frac{P(x|\theta) \times P(\theta)}{P(x)}$$

El arqueólogo bayesiano y la incertidumbre de las hipótesis

Como se describió anteriormente, la inferencia bayesiana sobre Monico-2 proporcionada al reportero se basó únicamente en los nuevos datos del sitio y la experiencia previa del arqueólogo con Monico-1. Sin embargo, si este quiere dar al reportero la mejor estimación posible, podría usar toda la evidencia disponible, incluidos los datos de Monico-2, su conocimiento experto e información de otros sitios arqueológicos. Para hacer esto, el arqueólogo revisa la literatura publicada e identifica información adicional sobre la proporción de perros con marcas de matanza recuperados de 38 sitios de Monico previamente excavados. Posteriormente busca investigar la variabilidad del comportamiento de matanza de cánidos, como lo demuestra la proporción de perros con marcas de matanza en cada

sitio de Monico, con miras a obtener una declaración previa probabilística sobre el parámetro θ (la proporción de perros con marcas de matanza).

La **Tabla 5** ilustra la distribución de los valores de θ a través de la frecuencia y proporciones de los sitios. La tabla muestra que de los 38 sitios, 20 tenían entre 0 % y 5 % de perros con evidencia de marcas de matanza. Doce sitios tienen entre 6 % y 15 % de perros que muestran estas huellas, mientras que otros 4 sitios reportan valores de θ entre 16 % y 35 %. Mientras tanto, otros 2 sitios arqueológicos reportan que θ oscila entre 36 % y 75 %. No hay sitios con más del 75 % de restos de perros que muestren huellas de este comportamiento.

Para empezar, el arqueólogo habla con otros expertos sobre nutrición, arqueología de los alimentos, y arqueología y etnografía de Monico. Con base en su conocimiento científico, plantea como hipótesis que, para considerar que los perros son

TABLA 5. Distribución de frecuencias del número de sitios con proporciones informadas de restos de perros con marcas de matanza (θ) y la proporción del número total de sitios con marcas de matanza en huesos de perros (probabilidades previas)

Proporción de restos de perros con marcas de matanza (θ)	Número de sitios con θ reportado	Proporción del número total de sitios con (θ) reportado (probabilidad previa)
0.00–0.05	20	0.53
0.06–0.15	12	0.32
0.16–0.25	3	0.08
0.26–0.35	1	0.03
0.36–0.45	1	0.03
0.46–0.55	0	0.00
0.56–0.65	0	0.00
0.66–0.75	1	0.03
0.76–0.85	0	0.00
0.86–0.95	0	0.00
0.96–1.0	0	0.00
Total	38	1.00

una contribución alimentaria sustancial en un sitio de Monico, debería haber evidencia de marcas de matanza en al menos el 50 % de los individuos. "Entonces —piensa el arqueólogo— mi primera hipótesis, H_1 , es que el valor de θ debería ser al menos del 50 %, o 0.5, para cualquier sitio específico de Monico. ¿Cuál es la probabilidad de que esta hipótesis sea correcta para Monico-2 según los datos que tengo y mi conocimiento previo?"

La muestra del sitio Monico-2 indicó que, de 10 individuos, 9 tenían marcas de matanza (por lo tanto, $\theta = 0.9$). El arqueólogo quiere usar el conocimiento previo, incluida la información obtenida de la revisión de la literatura, para comprender la variabilidad de θ en los sitios de Monico.

Por ello comienza registrando las proporciones de matanza de perros (θ) de los 38 sitios encontrados en la literatura. Para resumir estos datos, en la **Tabla 5** (columna 1) agrupa los valores de θ en intervalos iguales en incrementos de 0.10 (10 %, excepto

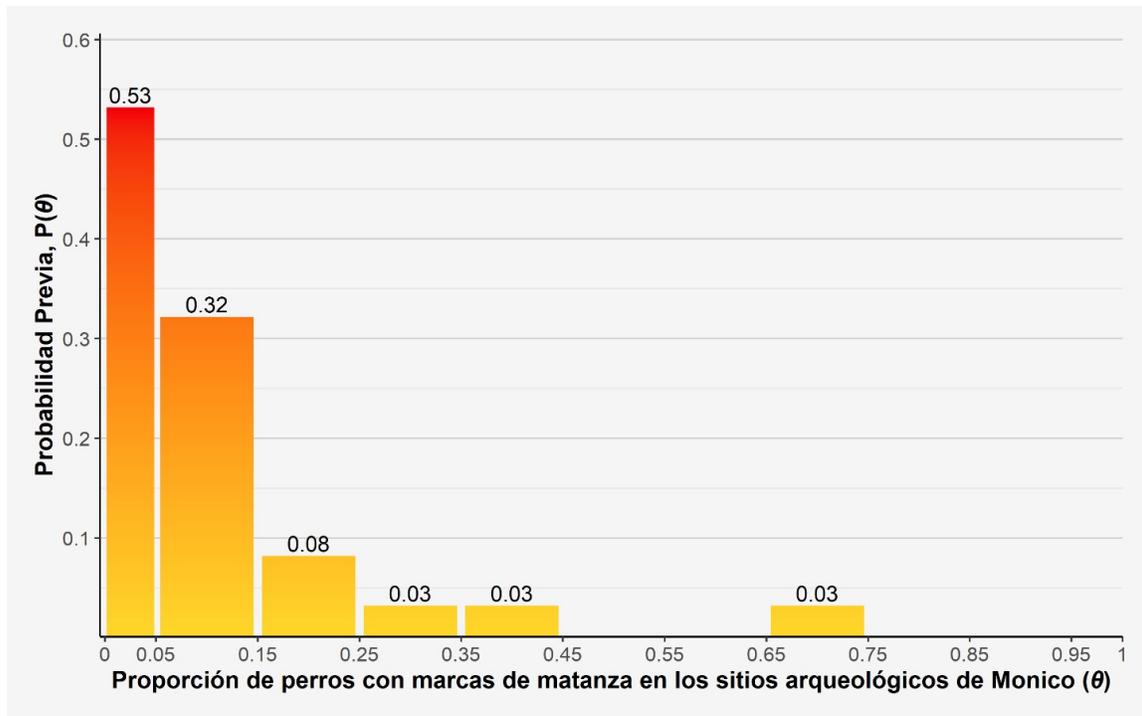


FIGURA 3. Representación simple de la distribución de las probabilidades previas del arqueólogo de las estimaciones de θ , la proporción de perros con marcas de matanza en los sitios arqueológicos de Monico (de la [Tabla 5](#)). Tenga en cuenta que los valores pequeños de θ tienen una probabilidad previa más alta que los más grandes.

el primer intervalo, que es menor). También registra el número de sitios que reportan valores de θ en cada intervalo (columna 2). Luego, el arqueólogo calcula la probabilidad previa de cada intervalo θ dividiendo el número en cada celda de la columna 2 por el número total de sitios, es decir, 38. De esta forma, la tercera columna reporta la proporción de sitios dentro de cada intervalo θ . Esta distribución de frecuencia también sirve como distribución previa de los valores de θ .

Luego, el arqueólogo traza la distribución de la proporción de perros sacrificados en los sitios de Monico ([Tabla 5](#)) para visualizar el conocimiento previo resultante que se puede derivar de este conjunto de datos ([Figura 3](#)).

Recuerde que en el marco bayesiano se necesita la verosimilitud ($P(x|\theta)$), la probabilidad de los datos ($P(x)$) y la probabilidad previa de la hipótesis ($P(\theta)$) para calcular la probabilidad posterior, es decir la probabilidad de que $\theta > 0.50$, dados los datos ($P(\theta > 0.5|x)$). La [Figura 3](#) ilustra la probabilidad previa, $P(\theta)$, para diferentes valores de θ .

Tenga en cuenta que, en contraste con los valores de un solo evento en los ejemplos anteriores, los componentes del teorema de Bayes en este caso son distribuciones de valores. La aplicación de estadísticas bayesianas en tales situaciones brinda una ventaja particular porque el marco permite a los arqueólogos evaluar la probabilidad de una hipótesis y la incertidumbre asociada. Por lo tanto, para continuar con el análisis bayesiano de los datos de Monico-2 a la luz del conocimiento previo de los 38 sitios (representados en la [Figura 3](#)), el arqueólogo necesita un modelo para representar la probabilidad de los datos, x , y parámetro(s) asociados, θ , para calcular la probabilidad, $P(x|\theta)$, y la probabilidad de los datos, $P(x)$.

La verosimilitud

Para calcular la probabilidad de los datos de Monico-2 dada la hipótesis, el arqueólogo necesita una función que pueda representar la verosimilitud, $P(x|\theta)$, de estos datos, x , dado el parámetro de interés, θ . Los arqueólogos emplean con frecuencia una función de probabilidad denominada modelo binomial para calcular la verosimilitud de datos compuestos por observaciones binarias, es decir, observaciones expresadas como 1/0, sí/no, verdadero/falso o presente/ausente. En este caso, el modelo binomial es apropiado para las observaciones que indican la presencia o ausencia de marcas de matanza en los esqueletos de perros, como en los datos de Monico-2. Como tal, el arqueólogo quiere calcular la verosimilitud de que 9 de cada 10 esqueletos de perros de este sitio exhiban marcas de matanza.

Matemáticamente, el modelo binomial se expresa por:

$$P(x|\theta) = \binom{N}{k} \times \theta^k \times (1 - \theta)^{N-k}.$$

Los símbolos k y N representan los datos: k es el número de perros observados con marcas de matanza, mientras que N es el total de perros observados. El parámetro del modelo, θ , en este ejemplo representa la proporción de perros con marcas de matanza de todos los perros observados en Monico-2.

El arqueólogo utiliza el método de estimación de parámetros llamado máxima verosimilitud (ML por sus siglas en inglés) para determinar el valor más probable de θ que habría generado los datos. La máxima verosimilitud pregunta, bajo el modelo binomial, ¿qué valor de θ es más probable que conduzca a

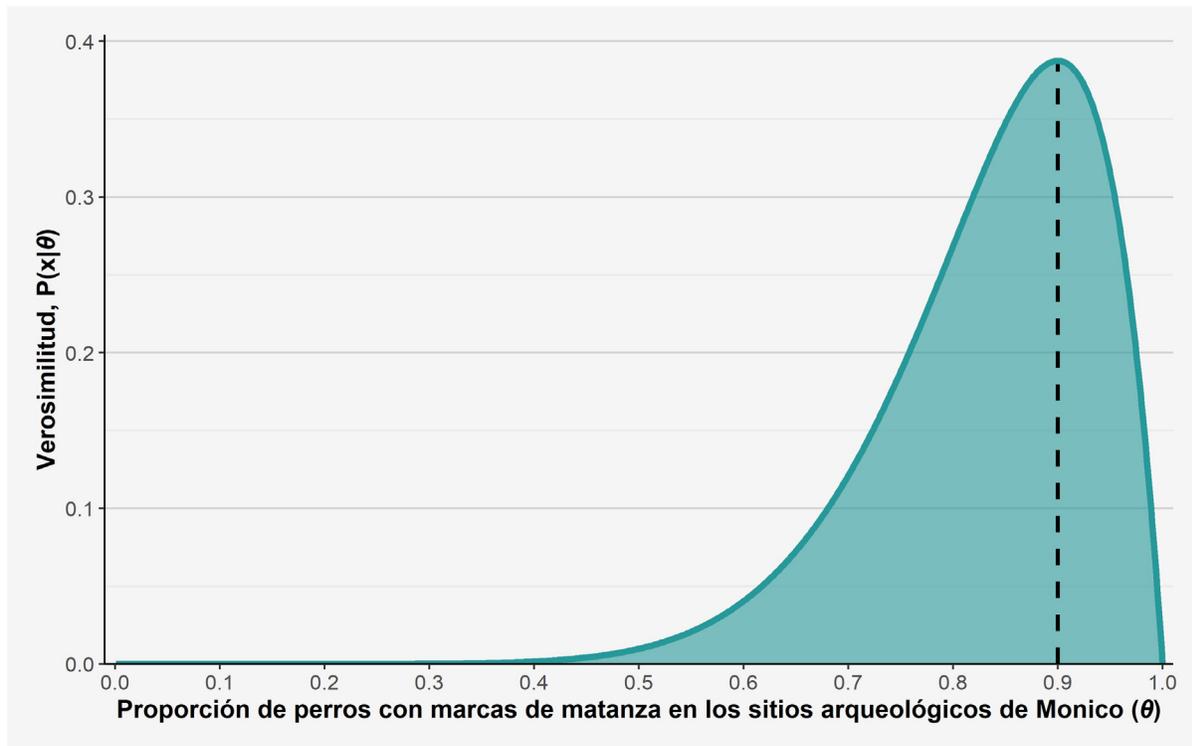


FIGURA 4. Distribución de valores de verosimilitud estandarizados correspondientes a cantidades variables de θ en el rango 0, 1. La línea negra discontinua indica el valor de θ que maximiza la verosimilitud de los datos. Esto es conocido como la estimación ML de θ .

los datos observados? En este caso, los datos binomiales del arqueólogo son $k = 9$ perros con marcas de matanza y $N = 10$ perros en total. La máxima verosimilitud evalúa qué valor del parámetro θ maximiza $P(x|\theta)$, la verosimilitud, en un rango sistemático de cantidades entre 0 y 1.

Para estimar el valor más probable de θ , el arqueólogo supone que la probabilidad de observar cada perro sacrificado es independiente de los demás, haciendo que la probabilidad de observar 9 perros sacrificados sea, θ^9 . Por el contrario, la probabilidad de observar un solo perro sin sacrificar es $(1-\theta)^{10-9}$, y la probabilidad de que haya 9 perros sin sacrificar y 1 perro sin sacrificar es $\theta^9 \times (1-\theta)^{10-9}$. Sin embargo, para calcular la verosimilitud de los datos, el arqueólogo también debe tener en cuenta el número de formas diferentes en que las 9 observaciones de perros con marcas de matanza, k , pueden ocurrir en la secuencia de 10 observaciones de perros, N .

El modelo binomial hace esto calculando $\binom{N}{k}$, conocido como

el coeficiente binomial (léase como “ N elige k ”). En este caso, si las identificaciones positivas de marcas de matanza en perros están representadas por 1 y las marcas de matanza no son 0, el coeficiente binomial calcula cuántos conjuntos desordenados podrían haber resultado en nueve 1 y un 0: por ejemplo, $x = \{0, 1, 1, 1, 1, 1, 1, 1, 1, 1\}, \{1, 1, 1, 1, 0, 1, 1, 1, 1, 1\}, \{1, 1, 1, 1, 1, 1, 1, 1, 1, 0\}, \dots$ etc². El coeficiente binomial es abreviado y puede calcularse usando la siguiente ecuación:

$$\binom{N}{k} = \frac{N!}{k! \times (N - k)!}$$

donde ! es la función factorial que produce el producto de un número entero y todos los números enteros debajo de él. En nuestro caso, $N=10$ y $k=9$, así que:

$$N! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3,628,800;$$

$$k! = 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 362,880;$$

$$\text{y, } (N - k)! = (10 - 9)! = 1! = 1.$$

Por lo tanto,

$$\binom{N}{k} = \frac{N!}{k! \times (N - k)!} = \frac{10!}{9! \times (10 - 9)!} = \frac{3,628,800}{362,880 \times 1} = 10.$$

Una vez que se ha calculado $\binom{N}{k}$, el arqueólogo puede continuar

estimando el valor de verosimilitud de una cantidad dada de θ calculando:

$$P(x|\theta) = 10 \times \theta^9 \times (1 - \theta)^{10-9}$$

en el rango de valores θ de 0 a 1 para encontrar la distribución de verosimilitud de los datos y, por lo tanto, el valor de θ que maximiza la función de verosimilitud. Este planteamiento se ilustra en la [Figura 4](#), de la cual el arqueólogo aprende que la estimación de ML de θ (dados los datos de Monico-2) es 0.9; en otras palabras, las observaciones en Monico-2 son más probables si la proporción de perros sacrificados en el sitio (θ) también es 0.9 (o 90 %).

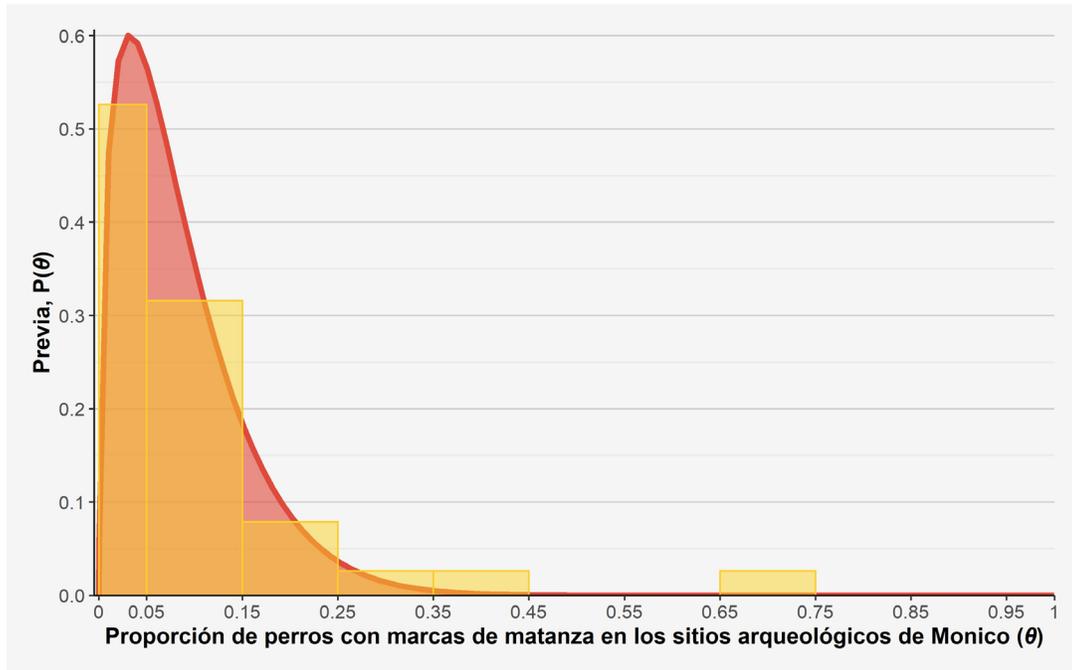


FIGURA 5. Modelo de probabilidad beta estandarizado, con parámetros $\alpha = 1.5$ y $\beta = 16$, que representan las probabilidades previas del arqueólogo representadas en la Figura 3. Tenga en cuenta la similitud con la Figura 3 en términos de forma y, en particular, la ubicación de la moda y el rango de valores.

La previa

Así como utiliza el modelo de probabilidad binomial para obtener la distribución de verosimilitud de los datos de Monico-2, el arqueólogo puede usar otro modelo de probabilidad para expresar $P(\theta)$, la distribución de probabilidad de θ , también conocida como previa. En este caso necesita una función de probabilidad que modele la distribución de θ , la proporción de perros con marcas de matanza, en los 38 sitios observados antes de la excavación de Monico-2. Los estadísticos usan con frecuencia la función de probabilidad beta para modelar la distribución de proporciones como θ . La expresión matemática del modelo beta es:

$$P(H) = P(\theta) = \theta^{\alpha-1} \times (1 - \theta)^{\beta-1}.$$

Por lo tanto, la forma del modelo beta está controlada por dos parámetros, a y b , que a su vez controlan las estadísticas de resumen clave, como la media y la varianza del modelo. A diferencia del modelo de verosimilitud, el arqueólogo en este caso quiere encontrar una distribución de θ que describa cuantitativamente su conocimiento previo. Para hacer esto, los parámetros beta pueden acomodarse para que se ajusten a la forma de la distribución de datos previa en la Figura 3. A través de un mejor ajuste visual, el arqueólogo estima que los valores $\alpha = 1.5$ y $\beta = 16$ dan como resultado una distribución de probabilidad que se asemeja a la del conocimiento previo sobre θ (es decir, la forma que se muestra en la Figura 3). Así, la distribución de la probabilidad,

$$P(H) = P(\theta) = \theta^{(1.5-1)} \times (1 - \theta)^{(16-1)},$$

en todos los valores de θ entre 0 y 1 se ilustra en la Figura 5.

La posterior

El arqueólogo es consciente de que los estadísticos utilizan con frecuencia las distribuciones binomial y beta en el contexto de los análisis bayesianos porque juntas funcionan bien para modelar las distribuciones de verosimilitud y de probabilidad previa, respectivamente, lo que simplifica los cálculos necesarios para calcular la posterior. Tales pares convenientes de modelos de probabilidad se conocen como *conjugadas*. Como resultado de las elecciones de modelado realizadas, el arqueólogo puede combinar algebraicamente los datos de verosimilitud binomial con los parámetros de la distribución beta previa para producir una distribución beta posterior representada por:

$$\begin{aligned} P(H|D) &= P(\theta|x) \\ &= \theta^{(k_{\text{verosimilitud}} + a_{\text{previa}} - 1)} \times (1 - \theta)^{(N_{\text{verosimilitud}} - k_{\text{verosimilitud}} + b_{\text{previa}} - 1)} \\ P(\theta|x) &= \theta^{(9 + 1.5 - 1)} \times (1 - \theta)^{(10 - 0 + 16 - 1)}. \end{aligned}$$

Por lo tanto, la verosimilitud y $P(\theta)$, las probabilidades previas, generan valores de $P(x|\theta)$, para calcular $P(\theta|x)$, la distribución de probabilidad posterior, a través de una cuadrícula fina de θ en el intervalo 0, 1 (1,000 valores entre 0 y 1). Estos están ilustrados en la Figura 6.

Después, el arqueólogo se centra en $P(\theta|x)$, la distribución posterior. El posterior le ayudará a hacer inferencias sobre la probabilidad de θ y su incertidumbre asociada (Figura 6). Así puede representar visualmente la estimación de θ (la proporción esperada de perros con marcas de matanza en Monico-2, según los datos observados y el conocimiento previo de los otros 38 sitios arqueológicos de Monico) y el rango de incertidumbre del 90 % de su estimación con un gráfico en la Figura 7.

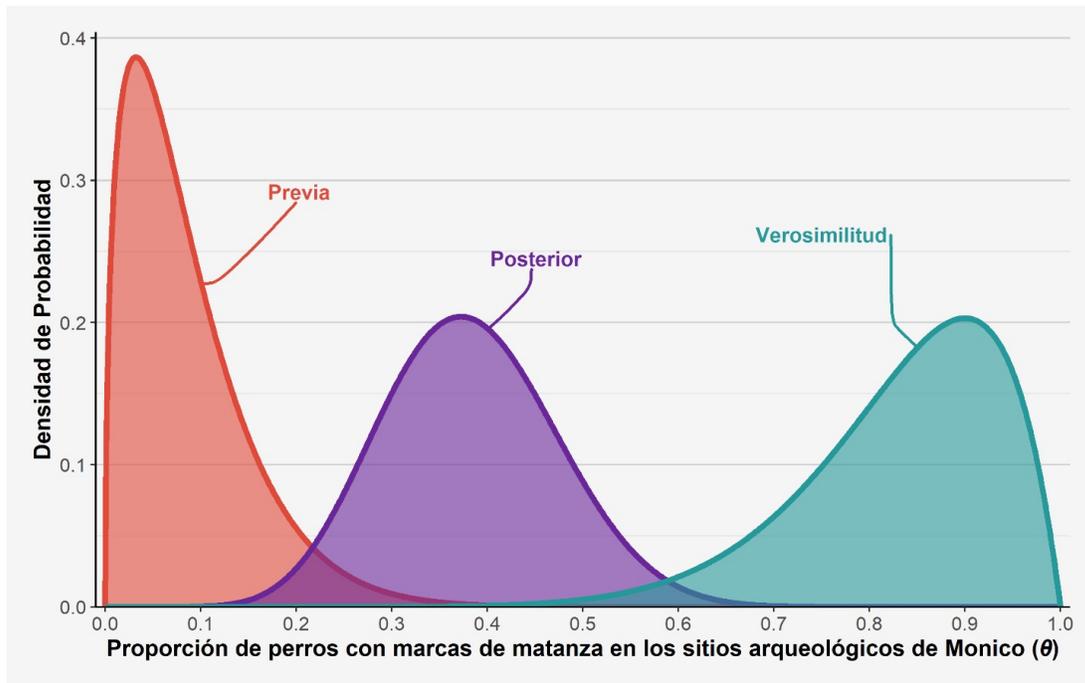


FIGURA 6. Distribuciones de las probabilidades previas del arqueólogo, la verosimilitud de los datos y las probabilidades posteriores. Todas las densidades de probabilidad están estandarizadas por una constante de normalización.

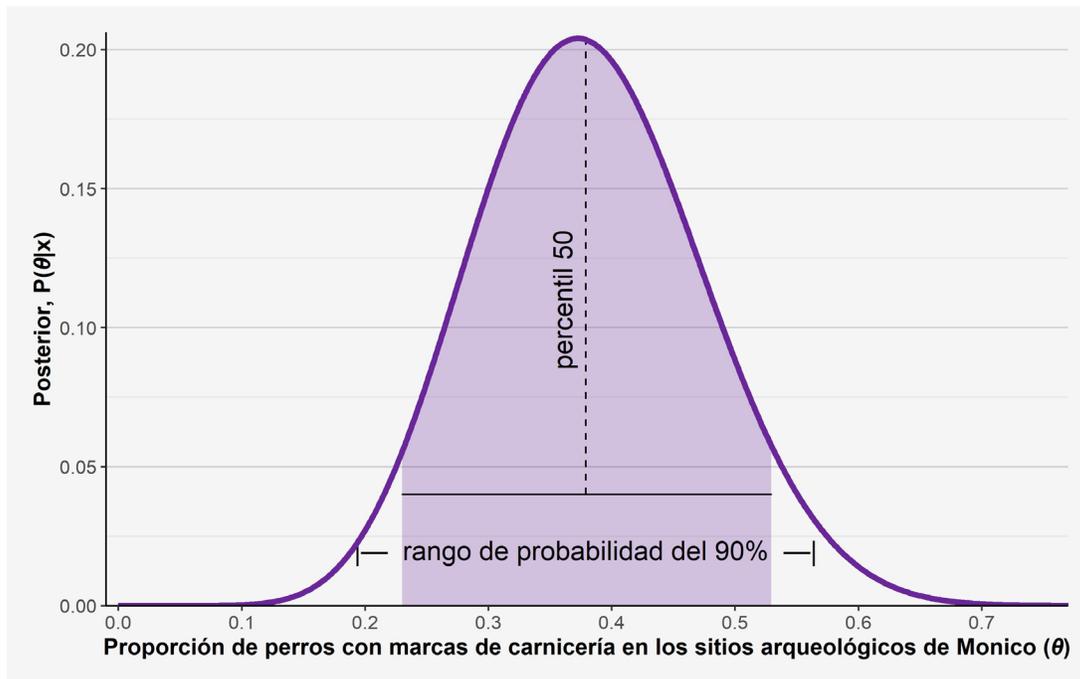


FIGURA 7. Distribución de probabilidad posterior con la línea punteada azul que muestra la estimación mediana (percentil 50) (0.38). La línea negra sólida representa el intervalo de densidad de probabilidad del 90 % (0.23-0.53).

A diferencia del marco de la NHST, la probabilidad posterior bayesiana permite al arqueólogo asignar probabilidades a las hipótesis sobre los valores de los parámetros. En este caso, la hipótesis es que el valor de θ , la proporción de perros

sacrificados en Monico-2, es superior a 0.5 (50 %, [Tabla 6](#)). Los valores que se muestran en la [Tabla 6](#) son inferencias resultantes de cálculos realizados utilizando la distribución posterior. El arqueólogo calculó la probabilidad de que θ sea mayor que 0.5

TABLA 6. Inferencias sobre θ a partir de la distribución de probabilidad posterior

P($\theta > 0.5$)	Valores de θ entre los percentiles 5 y 95		
	5%	50%	95%
0.10	0.24	0.38	0.54

(valor superior a la izquierda en la tabla) y los valores de θ en los percentiles de probabilidad 5, 50 y 95. Recuerde que junto con otros científicos propuso anteriormente que las marcas de corte deberían aparecer en al menos el 50 % (o 0.5) de los restos de perros en un sitio de Monico para considerar a los perros como "una importante contribución de alimentos". Sin embargo, la [Tabla 6](#) muestra que el valor de θ solamente tiene un 10 % de probabilidad de ser mayor al 50 %. Por lo tanto, la inferencia de que los perros eran una parte sustancial de la dieta en Monico-2 no es muy probable. Por ejemplo, el arqueólogo piensa: "Si un meteorólogo me dijera que hoy hay un 10 % de probabilidad de lluvia, no llevaría paraguas".

Es importante destacar que la incertidumbre en torno al valor de θ también puede expresarse como un intervalo de probabilidad. En el marco bayesiano, estos *intervalos de probabilidad se conocen como los intervalos de densidad de probabilidad más alta* (HPDI) y difieren de los IC de la NHST. Una de las diferencias más importantes es que la interpretación del HPDI es mucho más sencilla. El HPDI es la probabilidad del parámetro dados los datos, mientras que, como describimos anteriormente, el IC no es una probabilidad sobre el valor de la estimación del parámetro. En el caso de θ , la [Figura 7](#) le dice al arqueólogo que existe mucha incertidumbre en torno al verdadero valor de θ . Por ejemplo, la estimación de la mediana, o percentil 50, de θ es 0.38, lo que significa que, una vez que se incorpora la información previa disponible de la literatura y los datos de Monico-2, es muy probable que se observe que los ocupantes de Monico-2 incluyeron perros en su dieta el 38 % de las veces. Sin embargo, el HPDI del 90 % abarca de 0.23 a 0.53 (23 % a 53 %), lo que significa que, según nuestra información previa y los datos actuales, existe un 90 % de posibilidades de que θ se encuentre entre estos valores y solo un 10 % de posibilidades de que es mayor o menor que estos límites. Si bien la variación en θ alcanza más del 50 %, lo hace solo levemente y de nuevo es poco probable. Estos resultados significan que el arqueólogo no está muy seguro acerca de la propensión de los ocupantes a sacrificar perros (presuntamente) con fines dietéticos en Monico-2, especialmente considerando el pequeño tamaño de la muestra y el hecho de que los datos actuales de Monico-2 difieren bastante de aquellos en otros sitios.

CONCLUSIÓN

La inferencia bayesiana tiene ventajas para los arqueólogos que se extienden mucho más allá del ámbito de la calibración por radiocarbono y el modelado cronológico. La NHST fue muy útil para los arqueólogos durante muchos años, pero tiene limitaciones. Desafortunadamente, esa prueba basa la inferencia solo en datos nuevos debido a su estructura inherente. Su lenguaje y suposiciones pueden ser intrincados y confusos, y el

planteamiento no puede utilizarse para comparar directamente múltiples hipótesis de trabajo. La inferencia bayesiana supera muchos de estos problemas. Por otro lado, los arqueólogos a menudo analizan los problemas utilizando un marco bayesiano sin saber que lo están haciendo y sin recurrir a un marco probabilístico formal. Al igual que el arqueólogo bayesiano de nuestro caso ficticio, la mayoría de estos profesionales no realiza inferencias sobre el pasado empleando nuevos datos aislados del cuerpo de conocimiento existente. Por el contrario, actualizamos continuamente nuestro conocimiento previo con nueva evidencia para tomar decisiones, formar opiniones y generar conclusiones. La ventaja de la inferencia bayesiana sobre la NHST es que ofrece a los arqueólogos 1) un conjunto de herramientas más natural para aprender de los datos, 2) un lenguaje sencillo para hacer que las hipótesis sean cuantificables, explícitas y transparentes, y 3) la capacidad de utilizar la probabilidad para comparar múltiples hipótesis y realizar evaluación adicional.

En consecuencia, *el planteamiento bayesiano representa un cambio de paradigma en la inferencia arqueológica*. Las estadísticas bayesianas ofrecen un marco inferencial coherente que describe explícitamente la forma en que nuestra información previa se actualiza con nuevos datos para producir el estado actual del conocimiento. El proceso ayuda a evaluar el grado en que la evidencia actual y la nueva respaldan las hipótesis. Esto puede llevarse a cabo repetidamente hasta que haya una cantidad deseable de confianza (o falta de ella) en la precisión de una hipótesis. En este contexto, el marco bayesiano se asemeja a un proceso de aprendizaje similar a la investigación científica. Por ejemplo, los arqueólogos actualizan continuamente su conocimiento y grado de creencia en hipótesis utilizando nueva información obtenida a través de múltiples métodos de recopilación de datos, que incluyen excavación, estudio, experimentación, laboratorio y otras actividades analíticas.

Un número cada vez mayor de arqueólogos está utilizando estadísticas bayesianas para calibrar fechas de radiocarbono, construir cronologías y evaluar sus hipótesis sobre el pasado. La popularidad del *software* bayesiano relacionado con la cronología ha determinado que en ese contexto la inferencia bayesiana sea una operación simple, lo que significa que la mayoría de los usuarios encontrarán el *software* fácil de operar sin una comprensión básica de la lógica de la inferencia bayesiana y sus tres componentes fundamentales: la verosimilitud, la previa y la posterior. Además, sin esa comprensión fundamental, el poder analítico de las estadísticas bayesianas, más allá de la construcción de la cronología, puede no ser obvio, lo que ralentiza en lugar de mejorar su adopción más general.

Para mitigar este problema, este artículo destaca cómo los arqueólogos pueden utilizar la inferencia bayesiana para plantear cuestiones complejas a través de un ejemplo ficticio simple. Este planteamiento permite a los arqueólogos evaluar, comparar y actualizar sus hipótesis directamente, utilizando el peso de la evidencia y un proceso sencillo. Consideramos que este es uno de los impactos más significativos del paradigma bayesiano. Además, la inferencia bayesiana requiere que los arqueólogos sean conscientes y transparentes en relación a la información previa y actual para los análisis estadísticos dentro de una estructura probabilística. El marco incorpora explícitamente toda la información (previa y actual) para permitir una comprensión más completa de un problema.

Como resultado, las aplicaciones de este método conducen a la replicación, lo que permite que otros científicos arqueológicos las mejoren. Bajo esta luz, la inferencia bayesiana encaja con los esfuerzos en curso para promover métodos de ciencia abierta y datos abiertos en la investigación arqueológica. Este contexto anima a los investigadores a delinear todo el proceso lógico que subyace a sus resultados. Debido a sus ventajas, creemos que la inferencia bayesiana está bien posicionada para convertirse en un planteamiento estándar para evaluar hipótesis cuantitativas en arqueología.

Agradecimientos

Esta obra no necesitó permisos. Erik Otárola-Castillo agradece a Deb Nichols, John Watanabe, Sophie Nichols-Watanabe, Robert (Bob) L. Kelly y Dartmouth Coach por inspirar y facilitar el desarrollo de algunos conceptos en este artículo. Además, Amanda Veile, Mike Shott, Eduardo Fernández-Duque, Ben Marwick y dos revisores anónimos brindaron comentarios constructivos sobre borradores anteriores del manuscrito. Hannah Lipps nos brindó una ayuda invaluable con el formato de RMarkdown. Las magníficas ilustraciones de Warren Muzak (<http://www.warrenmuzak.com/>) permitieron que la cultura ficticia de Monico cobrara vida. Finalmente, Sarah Herr y el equipo editorial de AAP fueron muy gentiles y facilitaron el proceso de presentación y revisión por pares.

Declaración de disponibilidad de datos

Todos los archivos, el código R y los datos necesarios para replicar el manuscrito están disponibles como un documento R Markdown (.rmd) de Open Science Framework (DOI:10.17605/OSF.IO/54F62). Para mayor comodidad, el archivo R Markdown produce una preimpresión en PDF del manuscrito. Se puede acceder aquí: <https://osf.io/23bdt/>.

NOTAS

1. Este ejemplo se inspiró en creaciones similares en Thompson (1972), Flannery (1976, 1982), y los cuentos del valle de Uwasi (Shea 2020).
2. No se enumeran todos los conjuntos aquí, pero este ejemplo debería permitir al lector imaginar cómo puede ocurrir esto en un total de 10 maneras únicas. Aunque en este caso la solución es bastante simple, en otras aplicaciones podría no ser tan obvia, por ejemplo, el número de formas en que pueden ocurrir cinco éxitos en 10 intentos, es decir, 10 eligen 5 = 252.

REFERENCIAS CITADAS

- Aarts, Sil, Björn Winkens y Marjan van Den Akker
2012 The Insignificance of Statistical Significance. *European Journal of General Practice* 18: 50–52.
- Banning, Edward B.
2020 *The Archaeologist's Laboratory: The Analysis of Archaeological Evidence*. 2da ed. Springer, Cham, Suiza.
- Baxter, Michael J.
2003 *Statistics in Archaeology*. Arnold, Londres.
- Bayes, Thomas
1763 An Essay Towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions* 53: 370–418.
- Bayliss, Alex y Peter Marshall
2019 Confessions of a Serial Polygamist: The Reality of Radiocarbon Reproducibility in Archaeological Samples. *Radiocarbon* 61: 1143–1158.
- Bellhouse, David R.
2004 The Reverend Thomas Bayes, FRS: a Biography to Celebrate the Tercentenary of his Birth. *Statistical Science* 19: 3–43.
- Benjamin, Daniel J. y James O. Berger
2019 Three Recommendations for Improving the Use of *p*-Values. *The American Statistician* 73(sup 1): 186–191.
- Binford, Lewis R.
1964 A consideration of archaeological research design. *American Antiquity* 29: 425–441.
- Buck, Caitlin E., William G. Cavanagh y Cliff D. Litton
1996 *Bayesian Approach to Interpreting Archaeological Data*. Wiley, Chichester, Reino Unido.
- Buck, Caitlin E. y Bo Meson
2015 On Being a Good Bayesian. *World Archaeology* 47: 567–584.
- Carlson, David L.
2017 *Quantitative Methods in Archaeology Using R*. Cambridge University Press, Reino Unido.
- Chamberlin, Thomas C.
1965 The Method of Multiple Working Hypotheses. Reprinted. *Science* 148: 754–759. Publicado originalmente en 1890, *Science* (antigua serie) 15: 92–96.
- Chuard, Pierre J. C., Milan Vrtílek, Megan L. Head y Michael D. Jennions
2019 Evidence that Nonsignificant Results are Sometimes Preferred: Reverse *P*-Hacking or Selective Reporting? *PLOS Biology* 17:e3000127.
- Clarke, David L.
1968 *Analytical Archaeology*. Methuen, Londres.
- Cohen, Jacob
1994 The Earth is Round ($p < .05$). *American Psychologist* 49: 997–1003.
- Cowgill, George L.
1977 Trouble with Significance Tests and What We Can Do About It. *American Antiquity* 42: 350–368.
1993 Distinguished Lecture in Archeology: Beyond Criticizing New Archeology. *American Anthropologist* 95: 551–573.
2001 *Past, present, and future of quantitative methods in United States archaeology*. Stančić, Zoran y Tatjana Veljanovski (eds.), *Computing Archaeology for Understanding the Past. CAA 2000. Computer Applications and Quantitative Methods in Archaeology*, pp. 35–40. *BAR International Series* 931. *British Archaeological Reports, Oxford*.
- Diez, David M., Christopher D. Barr y Mine Çetinkaya-Rundel
2019 *OpenIntro Statistics*. 4ta ed. OpenIntro, Boston.
- Doran, James E. y Frank R. Hodson
1975 *Mathematics and Computers in Archaeology*. Harvard University Press, Cambridge, Massachussets.
- Drennan, Robert D.
2009 *Statistics for Archaeologists: A Commonsense Approach*. 2da ed. Springer, Boston.
- Dunson, David B. y James E. Johndrow
2020 The Hastings algorithm at fifty. *Biometrika* 107:1–23.
- Fisher, Daniel C.
1987 Mastodont procurement by Paleoindians of the Great Lakes region: hunting or scavenging? Nitecki, Matthew H. y Doris V. Nitecki (eds.), *The Evolution of Human Hunting*, pp. 309–421. Springer, Boston.
- Fisher, Ronald A.
1922 On the Interpretation of χ^2 from Contingency Tables, and the Calculation of *P*. *Journal of the Royal Statistical Society* 85: 87–94.
1925 *Statistical methods for research workers*. Oliver and Boyd, Edinburgo/Londres.
1935 *The Design of Experiments*. Oliver and Boyd, Edinburgo.

- Fletcher, Mike y Gary R. Lock
2005 *Digging Numbers: Elementary Statistics for Archaeologists*. 2da ed. Oxford University School of Archaeology, Reino Unido.
- Flannery, Kent V. (editor)
1976 *The Early Mesoamerican Village*. Academic Press, Nueva York.
- Flannery, Kent V.
1982 The Golden Marshalltown: A Parable for the Archeology of the 1980s. *American Anthropologist* 84: 265–278.
- Freeman, Peter R.
1976 A Bayesian Analysis of the Megalithic Yard. *Journal of the Royal Statistical Society: Series A (General)* 139: 20–35.
- Gelman, Andrew
2006 Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics* 48:432–435.
2018 The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Personality and Social Psychology Bulletin* 44: 16–23.
- Gelman, Andrew, Jennifer Hill y Masanao Yajima
2012 Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* 5: 189–211.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari y Donald B. Rubin
2020 *Bayesian Data Analysis*. 3rd ed. Chapman and Hall/CRC press, Boca Ratón, Florida.
- Gilks, Walter R., Sylvia Richardson y David Spiegelhalter
1996 *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC Press, Boca Ratón, Florida.
- Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman y Douglas G. Altman
2016 Statistical tests, p-Values, Confidence Intervals, and Power: a Guide to Misinterpretations. *European Journal of Epidemiology* 31: 337–350.
- Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn y Michael D. Jennions
2015 The Extent and Consequences of P-Hacking in Science. *PLOS Biology* 13: [e1002106](https://doi.org/10.1371/journal.pbio.1002106).
- Howson, Colin y Peter Urbach
2006 *Scientific Reasoning: the Bayesian Approach*. 3ra ed. Open Court Publishing, Chicago.
- Hubbard, Raymond
2011 The Widespread Misinterpretation of p-Values as Error Probabilities. *Journal of Applied Statistics* 38: 2617–2626.
- Hubbard, Raymond y María Jesús Bayarri
2003 Confusion over measures of evidence (p's) versus errors (α 's) in classical statistical testing. *The American Statistician*, 57: 171–178.
- Ioannidis, John P. A.
2005 Why Most Published Research Findings Are False. *PLOS Medicine* 2: [e124](https://doi.org/10.1371/journal.pmed.0020124).
- Jaynes, Edwin T.
2003 *Probability Theory: The Logic of Science*. Cambridge University Press, Reino Unido.
- Johnson, Douglas H.
1999 The Insignificance of Statistical Significance Testing. *The Journal of Wildlife Management* 63: 763–772.
- Kramer, Karen L, Amanda Veile y Erik R. Otárola-Castillo
2016 Sibling Competition & Growth Tradeoffs. Biological vs. Statistical Significance. *PLOS One* 11: [e0150126](https://doi.org/10.1371/journal.pone.0150126).
- Marwick, Ben
2017 *Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation*. *Journal of Archaeological Method and Theory* 24: 424–450.
- McCall, Grant S
2018 *Strategies for Quantitative Research: Archaeology by Numbers*. Routledge, Londres.
- McElreath, Richard
2020 *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2da ed. CRC press, Boca Ratón, Florida.
- McPherron, Shannon P., Will Archer, Erik R. Otárola-Castillo, Melissa G. Torquato y Trevor L. Keevil
2021 *Machine Learning, Bootstrapping, Null Models, and Why We Are Still Not 100% Sure Which Bone Surface Modifications Were Made by Crocodiles*. *Journal of Human Evolution* 164: [103071](https://doi.org/10.1016/j.jhevol.2021.103071).
- McShane, Blakeley B. y David Gal
2016 *Blinding Us to the Obvious? The Effect of Statistical Training on the Evaluation of Evidence*. *Management Science* 62(6): 1707–1718.
- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller y Edward Teller
1953 Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* 21: 1087–1092.
- Myers, Oliver H.
1950 *Some Applications of Statistics to Archaeology*. Egyptian Antiquities Service, El Cairo.
- Naylor, John C. y Adrian F.M. Smith
1988 An Archaeological Inference Problem. *Journal of the American Statistical Association* 83: 588–595.
- Neyman, Jerzy y Egon S. Pearson
1933 *On the Problem of the Most Efficient Tests of Statistical Hypotheses*. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231: 289–337.
- Otárola-Castillo, Erik R., Melissa G. Torquato y Caitlin E. Buck
2022 *The Bayesian Inferential Paradigm in Archaeology*. Pollard, Mark Ruth A. Armitage y Cheryl M. Makarewicz (eds.), *Handbook of Archaeological Sciences*, 2da ed. Wiley, Hoboken, Nueva Jersey.
- Otárola-Castillo, Erik R. y Melissa G. Torquato
2018 Bayesian Statistics in Archaeology. *Annual Review of Anthropology* 47: 435–453.
- Pearson, Karl
1900X. *On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling*. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50: 157–175.
- Robert, Christian y George Casella
2011 A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science* 26(1): 102–115.
- Salmon, Merrilee H.
1982 *Philosophy and Archaeology*. Academic Press, Nueva York.
- Shea, John J.
2020 *Prehistoric Stone Tools of Eastern Africa: A Guide*. Cambridge University Press, Nueva York.
- Shennan, Stephen
1997 *Quantifying Archaeology*. University of Iowa Press, Iowa.
- Spaulding, Albert C.
1953 Statistical Techniques for the Discovery of Artifact Types. *American Antiquity* 18: 305–313.
- Thiese, Matthew S., Zachary C. Arnold y Skyler D. Walker
2015 The Misuse and Abuse of Statistics in Biomedical Research. *Biochemia Medica* 25: 5–11.
- Thomas, David H.
1986 *Reguring Anthropology: First Principles of Probability and Statistics*. Waveland Press, Long Grove, Illinois.
- Thompson, Neil B.
1972 The Mysterious Fall of the Nacirema. *Natural History* 81: 412–417.
- Valeggia, Claudia R. y Eduardo Fernández-Duque
2022 Moving Biological Anthropology Research Beyond $p < 0.05$. *American Journal of Biological Anthropology* 177(2): 193-195.
- Vescelius, Gary S.
1960 *Archaeological Sampling: a Problem of Statistical Inference*. Gertrude E. Dole y Robert L. Carneiro (eds.), *Essays in the science of culture in Honor of Leslie White*. Thomas Y. Crowell, Nueva York.
- Vidgen, Bertie y Taha Yasseri
2016 P-values: Misunderstood and Misused. *Frontiers in*

Physics 4: 6.

Walker, John y Joseph Awange

2020 *Total Station: Measurements and Computations*. Walker, John y Joseph Awange (eds.), *Surveying for Civil and Mine Engineers: Acquire the Skills in Weeks*, pp. 77–99. Springer International, Cham, Suiza.

Wasserstein, Ronald L., Allen L. Schirm y Nicole A. Lazar

2019 *Moving to a World Beyond “ $p < 0.05$ ”*. *The American Statistician* 73(sup 1): 1-19, Londres.

Wolfhagen, Jesse

2019 *Rethinking Human-Cattle Interactions at Çatalhöyük (Turkey) through Bayesian Analysis of Cattle Biometry and Behavior*. Tesis doctoral, Department of Anthropology, Stony Brook University.

2020 Re-examining the Use of the LSI Technique in Zooarchaeology. *Journal of Archaeological Science* 123: [105254](#).

Wolverton, Steve J., Jonathan Dombrosky y R. Lee Lyman

2016 *Practical Significance: Ordinal Scale Data and Effect Size in Zooarchaeology*. *International Journal of Osteoarchaeology* 26: 255–265.