

HACIA EL USO SENSATO DE LA SIGNIFICACIÓN ESTADÍSTICA EN LA REVISTA CHILENA DE ORNITOLOGÍA

Towards the wise use of statistical significance in the Revista Chilena de Ornitología

RICARDO A. FIGUEROA^{1,2} & SERGIO ALVARADO^{2,3,4,5}

¹Editor Jefe

²Unión de Ornólogos de Chile (AvesChile).

Correspondencia: revistachilenaornitologia@aveschile.cl

³Editor Estadístico

⁴Programa de Bioestadística, Instituto de Salud Poblacional, Facultad de Medicina, Universidad de Chile.

⁵Laboratorio de Ecología de Vida Silvestre (LEVS), Facultad de Ciencias Forestales y de la Conservación de la Naturaleza, Universidad de Chile.

Correspondencia: salvarado@uchile.cl

... todo experimento existe sólo para dar a los hechos la oportunidad de refutar la hipótesis nula. Fisher (1935).

La estadística recibe a menudo mala fama porque es muy fácil hacer un mal uso de ella, involuntariamente o no. Sylvan Barnet & Hugo Bedau (1993).

... las posibilidades inferenciales de la estadística (ya sea bayesiana, frecuentista u otras) siguen siendo profundamente limitadas. Greenland & Poole (2012).

La estadística cumple un papel relevante en el proceso y progreso científico (Hampel 1997, Scheiner 2001). Al igual que otros grupos de investigadores, los ornitólogos usamos pruebas o modelos estadísticos para inferir patrones o procesos y establecer generalizaciones y predicciones a partir de los resultados de nuestra investigación (Buckland 1982, Fowler & Cohen 1996, Underhill 1999). Este proceso, conocido como *inferencia estadística*, permite profundizar nuestro conocimiento empírico y teórico acerca de cómo las aves responden a los cambios ecológicos y ambientales (James & McCulloch 1985, North & Byron 1985, North 1994, Underhill 1999). Sin embargo, el uso de los procedimientos estadísticos inferenciales requiere una sólida justificación teórica y metodológica. Por lo tanto, la evaluación específica de los análisis estadísti-

cos es una de las fases relevantes del proceso editorial en cualquier revista científica (Parker *et al.* 2018).

En nuestra experiencia, los análisis estadísticos defectuosos están entre las principales debilidades de muchos de los manuscritos enviados a las revistas ornitológicas. Los mayores defectos de esos análisis es que son incorrectos, injustificados, sobredimensionados o incluso innecesarios. Por otra parte, muchos revisores afines con el tema central de un manuscrito no necesariamente son conocedores de las teorías estadísticas. Por lo tanto, rara vez sus revisiones incluirán comentarios respecto de la validez de los procedimientos estadísticos. Por otra parte, algunos revisores entrenados estadísticamente difieren respecto de la validez de los análisis estadísticos presentados en el mismo manuscrito. Así, algunas veces es problemático decidir qué recomendación estadística hacer a los autores para que refinen sus conclusiones a partir de su información.

Para fortalecer la revisión de los análisis estadísticos y hacer la mejor recomendación a los autores, nuestra revista incluye desde la edición de junio de 2021 una nueva figura editorial: el *editor estadístico*. El propósito detrás de esta figura es fortalecer nuestro equipo editorial desde el punto de vista estadístico y metodológico. El rol principal del editor estadístico es verificar si los procedimientos estadísticos presentados por los autores tienen validez a la luz de su diseño de muestreo y la naturaleza de la

información. El editor estadístico revisará los manuscritos de manera previa o simultánea a la revisión que hagan los colegas pares de tal manera que el editor jefe reciba una evaluación más informada. Además, el editor estadístico estará a cargo de la sección de *alfabetización estadística* que incorporaremos en los próximos números de nuestra revista. Todo lo anterior beneficiará a los autores ya que evitará que sus manuscritos estén sustentados en análisis estadísticos deficientes.

Dada la crisis actual de replicación experimental, la pérdida de confianza en el paradigma frecuentista y la “guerra estadística” entre frecuentistas y bayesianos (Gelman & Loken 2014, Wasserstein & Lazar 2016, Amrhein *et al.* 2017) es más necesario que nunca contar con asesoría estadística profesional. Además, muchos autores están cometiendo serios “pecados estadísticos” debido a la obsesión por los hallazgos significativos (quizá persiguiendo la fama) o simplemente para publicar artículos más rápidamente y parecer altamente productivos (Comroy 2019a, Makin & Orban de Xivry 2019, Orban de Xivry en Comroy 2019b). Así, consideramos imprescindible que los autores estén advertidos de estas situaciones para que decidan informada y reflexivamente cómo abordar sus análisis estadísticos. También es fundamental que los autores, revisores y editores pongan los principios filosóficos de la estadística por sobre los algoritmos y tecnificación de la estadística. Los editores de revistas ornitológicas estamos en una posición clave para promover cambios que conduzcan a usar la estadística sensatamente. Esto requiere que los autores y revisores tengan claro cuál es el sentido fundamental de la estadística y cuáles son las virtudes y los defectos de los paradigmas estadísticos existentes (Underhill 1999, Colling & Szűcs 2021, Sedgwick 2022).

¿Cómo los ornitólogos usamos la estadística?

Los ornitólogos usamos la estadística de manera descriptiva y de manera inferencial. En el primer caso, analizamos y resumimos información mediante tablas o gráficos de distribución de frecuencias, medidas de localización o tendencia central (*e.g.*, media, mediana, moda) y medidas de variabilidad (*e.g.*, varianza, desviación estándar). Tales descriptores estadísticos son útiles para establecer, por ejemplo, diferencias de tamaño corporal (*e.g.*, Egli 1996) o tendencia en la riqueza o abundancia de especies. Aunque los procedimientos estadísticos descriptivos no requieren supuestos teóricos ni dependen necesariamente de diseños experimentales, son esenciales para avanzar hacia los procedimientos inferenciales.

En el caso de la estadística inferencial, ésta es fundamental para decidir si las predicciones de nuestras hipótesis son admisibles o no (Gotelli & Ellison 2004, León-Guerrero & Frankfort-Nachmias 2018). Los proce-

dimientos inferenciales están condicionados fuertemente al uso riguroso del método científico. Antes de usar cualquier procedimiento estadístico inferencial debemos establecer una pregunta clara de investigación, una o varias hipótesis, articular sus predicciones, diseñar y ejecutar experimentos, y recolectar, organizar y resumir la información generada por nuestro diseño experimental (Quinn & Keough 2002, Gotelli & Ellison 2004). Mediante la estadística inferencial intentamos extraer conclusiones acerca de grandes grupos (poblaciones) a partir de pequeños grupos (muestras). Cuando hacemos esto, presumimos que las muestras son representativas de la población, que poseen naturaleza aleatoria y que nuestros datos son confiables.

Actualmente, los ornitólogos seguimos tres paradigmas estadísticos inferenciales: la estadística clásica o frecuentista, la estadística bayesiana y la selección de modelos basada en la teoría de la información. El paradigma frecuentista se sustenta en los conceptos clásicos de probabilidad, contraste de hipótesis (hipótesis nula [H_0] vs hipótesis alternativa [H_1]), pruebas de significancia, el nivel de significación o umbral alfa (α) y el valor de probabilidad (valor de p) (Fisher 1935). Dentro de un marco de estudio definido, los investigadores rechazan o no rechazan una hipótesis basándose en un nivel de significancia estadística (Fisher 1935). El término frecuentista alude a la idea de la frecuencia a largo plazo de los resultados en repeticiones infinitas de experimentos o muestras (Johnson 1999).

Los ornitólogos que se rigen por los procedimientos de la estadística bayesiana miden el efecto de un tratamiento incorporando la información previa de un determinado suceso para calcular su probabilidad posterior. Ellos cuantifican la fuerza de la evidencia basándose en el factor de Bayes. Mientras mayor sea el factor de Bayes mayor es la fuerza de la evidencia. Los ornitólogos que siguen esta línea determinan previamente esa probabilidad ya sea de manera subjetiva (*e.g.*, “creencia fundamentada”) u objetiva (*e.g.*, mediciones previas) (van de Schoot *et al.* 2021). Los ornitólogos que se rigen por la “teoría de la información” seleccionan entre varios modelos candidatos aquellos que contengan las variables con mayor poder explicativo mediante criterios de selección que minimizan la pérdida de información (Burnham & Anderson 2001). Ya que cada variable en un modelo representa a una hipótesis, la selección de modelos es congruente con el concepto de hipótesis múltiples (Chamberlin 1890).

Entre los paradigmas estadísticos descritos, la estadística frecuentista es prevalente en ornitología. La razón de esto es que la mayoría de los científicos y académicos han aceptado, usado y promovido este paradigma durante más de cinco décadas. Hasta hace algunos años, el enfoque

frecuentista fue el único enseñado en las universidades y el único explicado ampliamente en decenas de libros. Por lo tanto, muchas generaciones de ornitólogos egresamos de nuestras carreras con una “mente esencialmente frecuentista”. Aunque el paradigma frecuentista tiene varias virtudes, éste sufre de una crisis de confianza debido a los cuestionamientos sobre la utilidad inferencial de las pruebas de significación estadística. Sin embargo, las causas de esa crisis resultaron principalmente del mal uso y abuso de las pruebas de significancia y de la mala interpretación del valor de p . En parte, esto es una consecuencia de la enseñanza deficiente de la estadística en las universidades.

Aquí proporcionamos algunas sugerencias que orientará a los autores acerca de cómo abordar sensatamente el uso de la estadística inferencial antes de enviar sus manuscritos a nuestra revista. Aunque nos enfocamos principalmente en el paradigma frecuentista, muchas de nuestras sugerencias también son orientadoras para quienes miren hacia los otros paradigmas estadísticos. Sabemos que muchos autores son bien conocedores de los recursos estadísticos que son necesarios para su investigación, pero muchos otros sólo tienen una noción incipiente sobre el uso de la estadística. Así, nuestras sugerencias están dirigidas principalmente a quienes tienen poca o ninguna experiencia con el uso de la estadística inferencial, y a quienes la usan solo como un ritual dentro del ámbito de su investigación. En la elaboración de nuestras sugerencias seguimos esencialmente la línea reflexiva de Peter Feinsinger (Feinsinger 2001) y los fundamentos de la inferencia estadística según Ronald Fisher (Fisher 1922, 1956). Estas sugerencias las enriquecimos con varias reflexiones de otros autores.

Al final de esta editorial incluimos un glosario con varios de los términos técnicos y filosóficos que usamos a lo largo de nuestro relato. Aunque estos términos son elementales, no siempre los entendemos tan claramente, y algunos términos filosóficos son difíciles de definir y asimilar satisfactoriamente.

Estimado autor, aquí están nuestras sugerencias:

1. El uso apropiado de cualquier procedimiento estadístico inferencial requiere que tus resultados provengan de un estudio basado en un diseño experimental robusto. Asegura que tus unidades de estudio sean independientes (*i.e.*, que ninguna influya o afecte a las observaciones registradas en la otra), que estén distribuidas de manera aleatoria y sean seleccionadas aleatoriamente (James & McCulloch 1985, Kamil 1988). En muchas situaciones de campo no es posible garantizar la condición de independencia ni asignar aleatoriamente las unidades de estudio. Por lo tanto, debes ser cauteloso de tus conclusiones y declarar esa limitación en tu manuscrito (ver sugerencias 20). Idealmente, debes obtener un tamaño de muestra apropiado para responder satisfactoriamente tu pregunta de investigación u obtener apoyo a tu hipótesis biológica (ver Recomendación 3). El cálculo del tamaño de la muestra debiese considerar al menos algún criterio empírico (*e.g.*, ≥ 20 réplicas; Simmons et al. 2013). Sin estos requisitos esenciales, la aplicación de cualquier técnica o método estadístico inferencial te conducirá a resultados espurios o distorsionados. Existe una oferta variada de libros sobre diseño experimental. Para el diseño de estudios de campo recomendamos los libros de Feinsinger (2001) y Scheiner & Gurevitch (2001), y el capítulo sobre diseño de experimentos en el libro de Gotelli & Ellison (2004).
2. Cuando no puedas obtener un número estadísticamente aceptable de réplicas o unidades de estudios puedes seguir la “regla de 10” (Gotelli & Ellison 2004). Esta regla no tiene base teórica, pero refleja la experiencia ganada en el campo con diseños experimentales exitosos y no exitosos. En casos excepcionales (*e.g.*, una especie naturalmente rara), un mínimo de cinco unidades de muestreo (*e.g.*, sitio nido) por condición (*e.g.*, tipo de hábitat) podría ser aceptable. Existen pruebas estadísticas *ad hoc* para estas excepciones. Sin embargo, tu poder estadístico (ver sugerencia 3) será muy bajo y corres el riesgo que tus inferencias sean muy limitadas. Evita inferencias basadas en el valor de p si tus resultados provienen de muestras de tamaño pequeño, particularmente si esperas que el tamaño del efecto (ver Glosario) sea débil o moderado (Parker *et al.* 2018).
3. Considera medir el poder estadístico antes de diseñar tu estudio (ver Glosario). Al momento de calcular el tamaño de la muestra debes considerar en la fórmula cuál es el tamaño del error de tipo II de tu estudio (ver Glosario). El poder estadístico te informa si el número de unidades de estudio que determinaste fueron las suficientes para detectar un efecto, cuando ese efecto es real (Steidl & Thomas 2001, Lecoutre *et al.* 2005). Mientras mayor es el poder estadístico mayor es la posibilidad de detectar un efecto real. Sin embargo, ten en cuenta que un alto poder estadístico no siempre garantiza el hallazgo de un efecto real (Button *et al.* 2013). Nosotros recomendamos calcular el poder estadístico una vez levantada la muestra más bien para que estés informado de las limitaciones inferenciales de tu estudio. En general, los estudios de campo logran un bajo a mediano poder estadístico debido a la

dificultad de obtener suficientes réplicas (Halsey *et al.* 2015, Parker *et al.* 2018). Conocer el poder estadístico de tu estudio también te orientará respecto de la robustez de tu diseño experimental ante la posibilidad de usar métodos estadísticos no frecuentistas (Steidl & Thomas 2001).

4. Como ya mencionamos, en algunas ocasiones podrías tener dificultades para obtener un número estadísticamente apropiado de unidades de estudio. En estos casos es útil visualizar gráficamente tus datos e identificar cualitativamente alguna tendencia o asociación (Ellison 2001). También puedes describir tu información usando medidas de posición (*e.g.*, media, mediana, moda), de dispersión (*e.g.*, rango, desviación estándar, coeficiente de variación), de forma (*e.g.*, asimetría y curtosis) y análisis gráfico sencillo (*e.g.*, gráficos de barras, cajas o “box-plot” o histogramas). Este conjunto de procedimientos simples (conocido como análisis exploratorio de datos) debe preceder a cualquier análisis de mayor complejidad estadística. Existen varias pruebas estadísticas frecuentistas diseñadas para analizar muestras de tamaño pequeño (Zar 1999). Sin embargo, antes que decidas usarlas verifica si son válidas para el tipo de experimento en que basarás tu estudio.
5. Planifica tu análisis estadístico antes de implementar tu estudio y usa el procedimiento inferencial más simple que ajuste con tu diseño experimental. Si tus resultados provienen de estudios de campo, entonces prefiere pruebas o modelos no paramétricos. En general, los procedimientos no paramétricos son apropiados para muestras de tamaño pequeño y tienen menos restricciones que sus análogos paramétricos (Fowler & Cohen 1994, Siegel & Castellan 1995).
6. Ten siempre a mano textos estadísticos apropiados para tu área de estudio. En el caso de los diseños experimentales sencillos recomendamos el libro de Fowler & Cohen (1994). Dos libros bastante didácticos y entretenidos son los de Van Emden (2008) y Salkind (2017). Siempre asegúrate que los textos contengan descripciones de un amplio rango de pruebas estadísticas no paramétricas; te recomendamos el libro de Zar (1999) y el de Siegel & Castellan (1995). Sokal & Rohlf (1969) proporcionan descripciones bastante didácticas de algunos procedimientos no paramétricos. Todos los libros anteriores tienen un enfoque puramente frecuentista. Un texto que contiene también descripciones de procedimientos no frecuentistas es el de Quinn & Keough (2002). Dos buenos libros para para abrir la mente bayesiana son los de McElreath (2016) y de Field (2016). La mayoría de estos libros están escritos en inglés, pero hay varios libros de estadística básica escritos en español que están disponibles en la internet (*e.g.*, Blair & Taylor 2008).
7. Reflexiona siempre si necesitas realmente aplicar algún procedimiento estadístico inferencial, en particular las pruebas de significación. Aunque tu diseño experimental sea sólido y tu muestra sea grande, muchas veces las pruebas o modelos estadísticos son innecesarios o incluso inapropiados para analizar tus resultados (Sokal & Rohlf 1969, Cherry 1998, Johnson 1999). Muchos experimentos de campo quedan sujetos a variables ambientales no consideradas por el investigador. Si no tienes el control de tales variables es mejor que no apliques ningún procedimiento inferencial. Además, por razones logísticas o limitaciones naturales (*e.g.*, escasez natural de un cierto tipo hábitat) es posible que no puedas cumplir con el supuesto de independencia espacial de tus unidades de estudio. ¡Mantén la calma! En estos casos, podrías obtener una enorme cantidad de información biológica cuyo valor no aumentará con un análisis estadístico inferencial; de hecho, sería inapropiado. Tu diseño experimental puede ser indiscutiblemente robusto, pero si el tamaño del efecto (*e.g.*, magnitud de la diferencia) es evidente a simple vista no necesitas “engalanar” tu hallazgo con ningún “algoritmo estadístico”. Nuevamente, un despliegue gráfico de tus resultados combinado con recursos estadísticos descriptivos puede ser el camino más acertado para extraer lo mejor de tu información (Cherry 1998, Lang *et al.* 1998, Ellison 2001, Cumming 2012).
8. Antes de usar los procedimientos basados en pruebas de significación estadística, ten en cuenta que tales pruebas poseen varias limitaciones que las hacen inapropiadas o inútiles para estudios observacionales (Buckland 1982, Cherry 1998, Johnson 1999, Greenland & Poole 2013). En general, las pruebas de significación estadística son poco informativas y lógicamente pobres, el nivel α es arbitrario y sin base teórica, y el valor de p es vulnerable a interpretaciones erróneas (*e.g.*, Yoccoz 1991, Anderson *et al.* 2000, Martínez-Abraín & Oro 2005, Greenland & Poole 2013). Además, el valor de p conduce a decisiones dicotómicas en cuanto a admitir o no admitir la hipótesis de nulidad y no permite cuantificar la dirección de la diferencia encontrada. Finalmente, el valor de p es sensible al tamaño de la muestra (Underhill 1999). Por ejemplo, si dos ornitólogos conducen el mismo experimento

con diferentes tamaños de muestra, digamos 10 vs 30, ellos podrían llegar a conclusiones contrapuestas (*i.e.*, $p > 0,05$ vs $p < 0,05$).

9. Cuando uses justificadamente las pruebas de significación estadística, no etiquetes tus hallazgos usando los términos “estadísticamente significativo” (*e.g.*, $p < 0,05$) o “estadísticamente no significativo” (*e.g.*, $p > 0,05$). En el lenguaje común, el término “significativo” es interpretable como fiable. Sin embargo, un valor de $p < 0,05$ y un valor de $p > 0,05$ no son necesariamente una evidencia de que la hipótesis de nulidad es falsa o verdadera, respectivamente (Amrhein *et al.* 2019) ¡Recuerda que α es un valor arbitrario!
10. Evita la “significante-itis”; *i.e.*, la obsesión compulsiva por los hallazgos estadísticamente significativos (Chía 1997). Muchos autores que viven en la “cultura del valor de p ” (Nelder 1999) persiguen resultados estadísticamente significativos debido a que (i) confunden el valor de p como un indicador de la fuerza de la evidencia, y (ii) es la única manera que sus manuscritos sean aceptados por editores y revisores con la misma confusión (Underhill 1999, Martínez-Abraín & Oro 2005, Sedgwick 2022). La “cultura del valor de p ” puede conducir incluso a algunos investigadores a piratear el valor de p (“ p -hacking”); *i.e.*, a empujar sus valores p hacia límites significativos para que sus hallazgos parezcan relevantes y publicables (Comroy 2019a). Sin embargo, la significación estadística no refleja necesariamente la relevancia biológica o clínica de nuestros hallazgos (Potish *et al.* 1980, Krebs 1989, Yoccoz 1991, Underhill 1999, Malay 2016). Diferencias estadísticamente pequeñas pueden tener consecuencias biológicamente considerables. Siempre es reorientador recordar que la intención de Fisher fue que usáramos la significancia estadística solo como una herramienta para indicar que nuestros resultados justificaban más investigación (Fisher 1935, Sedgwick *et al.* 2022).
11. Cuando sea apropiado, construye los intervalos de confianza alrededor de la medida de interés (*e.g.*, media, proporción) para los distintos niveles o grupos de muestras (Martínez-Abraín & Oro 2005). Los intervalos de confianza te permiten saber que tan buenas son tus estimaciones, comparar confiablemente las estimaciones provenientes de distintos grupos de muestras y tener alguna idea de si hay algún efecto biológicamente relevante (*i.e.*, que tan grande es la magnitud del efecto). En la mayoría de los estudios, la estimación será más importante que la aplicación de una prueba estadística (Yoccoz 1991, Cherry 1998, Krebs 2000). Compara siempre de manera gráfica tus intervalos de confianza para visualizar la magnitud del efecto. Esto último aumenta el poder informativo de tus análisis (Cherry 1998, Johnson 1999, Ellison 2001).
12. Consulta a un bioestadístico profesional cuando no entiendas o tengas dudas sobre el uso de algún procedimiento estadístico complejo (Buckland 1980, Gustavii 2008). Si tu estudio es observacional, busca a un bioestadístico que tenga experiencia con diseños observacionales y conozca las limitaciones de la información proveniente de estudios conducidos en el campo. Ten en cuenta que algunos bioestadísticos pueden discrepar respecto del uso de ciertos procedimientos estadísticos lo cual puede confundirte aún más. Por ejemplo, un profesional puede exigirte que uses modelación estadística para analizar tus resultados obtenidos en apenas cinco sitios de muestreo, pero otro puede sugerirte simplemente que evalúes tu información con estadística descriptiva. En estos casos, cultiva tu sensatez y opta por la solución más simple y a la vez más informativa.
13. No uses procedimientos estadísticos sofisticados o novedosos solo para sorprender a los revisores o demostrar que tu estudio es estadísticamente “avanzado”, sobre todo si tienes pocas observaciones o tu estudio es meramente observacional. Ya que los análisis complejos y sofisticados “procesan” múltiples preguntas o múltiples hipótesis a la vez, los investigadores deben cumplir con varios supuestos teóricos y empíricos. Además, los usuarios de estos análisis deben realizar varias pruebas y ajustes poco explícitos antes de inferir algo de su información. Esto último conduce a un problema serio. Con cada prueba adicional aumenta la probabilidad que un investigador concluya erróneamente que existe al menos un efecto “estadísticamente significativo” (Gelman *et al.* 2012, Gelman & Loken 2014). Cuando los investigadores usan procedimientos complejos y sofisticados para extraer inferencias a partir de estudios exploratorios u observacionales sus conclusiones son a menudo espurias (James & McCulloch 1990).
14. Cuando uses justificadamente procedimientos estadísticos complejos y sofisticados, intenta describirlos de la manera más clara, simple, didáctica e intuitiva posible (*e.g.*, Fowler & Cohen 1996, Gustavii 2008, In & Lee 2017, Gelman 2018). Relata tus resultados pensando en el “cerebro izquierdo” y “cerebro derecho” de los lectores; *i.e.*, no solo entregues valores

numéricos para apoyar tus conclusiones, sino también ilumina a tus lectores relatando de manera simple la interpretación y el alcance de esos valores (Buckland 1980, Gelman 2018). Nuestra revista está enfocada hacia una amplia audiencia. Así, el universo de nuestros lectores puede tener un nivel de conocimiento estadístico ampliamente variable. Anderson *et al.* (2001) y Brennan (2012) ofrecen varias pautas para presentar de manera informativa los resultados de los análisis estadísticos dentro de un manuscrito. Cabe recordar que hace 36 años, John Gerrard (1985) afirmó que *la interacción entre ornitólogos amateurs y ornitólogos profesionales es una de las fortalezas de la ornitología respecto de otras áreas científicas*. Sin embargo, él también advirtió que esa fortaleza estaba en riesgo debido a que los ornitólogos profesionales estaban usando técnicas estadísticas sofisticadas incomprensibles para los ornitólogos amateurs. Sin duda, esa advertencia sigue vigente.

15. No asumas que los programas estadísticos, por sofisticados que sean, proporcionan resultados realmente confiables (*e.g.*, Eklund *et al.* 2016). Los programas computacionales nos permiten ahorrar tiempo y minimizar el trabajo a mano, pero son falibles (Littlewood & Strigin 1992). Además, los programas estadísticos no discriminan si tus datos provienen de un diseño experimental robusto o defectuoso. No confundas una librería o módulo de un paquete estadístico con un método estadístico. Por otra parte, los programas diseñados para la simulación de modelos pueden generar valores exacerbados de p que pueden conducirte a conclusiones espurias (White *et al.* 2014). Si tus resultados te parecen “sospechosos” o sin sentido biológico, entonces replantea tu análisis o busca la orientación de un experto. Mantente siempre dispuesto a repensar tus datos y reconsiderar los resultados de tus análisis (Carraway 2009). Siempre será revelador contrastar los indicadores estadísticos que proporciona tu programa computacional con una representación gráfica de tus resultados. Si encuentras que son congruentes, entonces quédate tranquilo. Si no, nuevamente cultiva tu sensatez y opta por la alternativa más simple y a la vez más informativa.
16. Mantente cauteloso ante la “propaganda” sobre las virtudes superiores de los paradigmas estadísticos emergentes. Varios investigadores han hecho un intenso y amplio llamado en la literatura a reemplazar la estadística frecuentista por la estadística bayesiana o por la selección de modelos (*e.g.*, Anderson *et al.* 2000, Anderson & Burnham 2002, Ellison 2004). Aunque estos mismos investigadores llaman a usar apropiadamente los procedimientos basados en esos paradigmas estadísticos emergentes, muchos otros tienden a entenderlos o usarlos equivocadamente. Si deseas mirar hacia esos paradigmas emergentes, te recomendamos que te informes bastante bien sobre sus bases teóricas y metodológicas. Si quieres comprender la inferencia bayesiana o selección de modelos es necesario que comprendas bien la inferencia frecuentista. En esos casos es estrictamente necesario que cuentes con la asesoría de un bioestadístico especializado.
17. No desprecies la filosofía. Usamos la estadística para decidir qué es verdad y qué no (Goodman 2016, Sedgwick 2022). Por lo tanto, necesitamos una comprensión básica de los fundamentos filosóficos de cada paradigma estadístico para decidir informadamente cuál seguiremos (Sedgwick 2022). Aunque parezca sorprendente, muchos ecólogos, incluyendo ornitólogos, no están conscientes de los paradigmas en los que operan (Krebs 2000). Conocer las bases filosóficas de cada paradigma estadístico también te orientará acerca del alcance de tus inferencias. La estadística nos auxilia en la búsqueda de la verdad contenida en la naturaleza. Sin embargo, los paradigmas estadísticos son constructos humanos y, por lo tanto, son falibles (Sedgwick 2022). Así, corremos el riesgo de que tales constructos nos conduzcan hacia un camino alejado de la verdad (Greenland & Poole 2013). Puede ser incómodo para ti dejar el teclado de tu computador y ya no “correr” tu programa estadístico “estrella”. Puede ser aburrido para ti leer sobre filosofía estadística. No obstante, no necesitas leer todo sobre Popper o Bayes; bastará una buena enciclopedia de filosofía estadística (*e.g.*, Zalta & Nodelman 2021). Si logras tener una pequeña noción sobre los fundamentos filosóficos de la estadística inferencial, podrás resistirte a los “cantos de sirena” de los paradigmas estadísticos emergentes.
18. Intenta educarte por ti mismo en estadística y aprende lo verdaderamente sustancial: los principios subyacentes de la inferencia estadística (Sedgwick 2022). Sabemos que los cursos universitarios de estadística están cargados hacia el uso mecánico de los “algoritmos estadísticos” (Sedgwick 2022). Esta práctica es acentuada por la oferta variada de programas computacionales. Ni hablar de los cursos de estadística de postgrados no estadísticos. Generalmente, los instructores de estos cursos “avanzados” agobian a sus estudiantes con decenas de artículos basados en procedimientos estadísticos complejos. Esto es lo que Salkind (2017) irónica y acertadamente llama “sadí-

tica". Por otra parte, algunos instructores confunden un curso de estadística con un curso sobre el uso de un programa computacional de análisis de datos. Ciertamente, la ausencia de pedagogía degrada seriamente la calidad de los cursos de estadística (Zieffler 2018, Sedgwick 2022). Afortunadamente hay muchos recursos en internet para aprender estadística de manera amigable, efectiva y autónoma. El portal de www.youtube.com es abundante en tutoriales para aprender conceptos y procedimientos estadísticos tanto básicos como complejos. Al respecto, recuerda siempre que primero debes dominar lo básico para avanzar hacia métodos complejos. Debido a que podrías encontrarte también con tutoriales poco explicativos, siempre contrasta la información con las explicaciones que están en los libros clásicos de estadística. Además, varios autores sensatos han escrito libros prácticos y atractivos sobre estadística (e.g., Fowler & Cohen 1994, Field 2016, Salkind 2017). Un libro bastante divertido enfocado a darle sentido y "sabor" a la estadística es el de Huff (1993).

19. Krebs (2000) nos aconsejó que dediquemos mucho más tiempo a enfocarnos en los temas ecológicos reales. Cuando estés en el campo reflexiona acerca de la razón de la inferencia estadística en ornitología. Recuerda que la inferencia estadística es el proceso de extraer conclusiones acerca de la población usando la información proveniente de un conjunto de muestras (a menudo pequeño) o incluso de una única muestra de esa misma población. Cuando visites tus unidades de muestreo en el campo, notarás que difícilmente obtendrás cada una de tus muestras exactamente bajo las mismas condiciones. La razón es que cada unidad de muestreo variará espacial y temporalmente, y porque nuestra percepción puede variar en cada unidad de muestreo. Esto último es difícil de controlar estadísticamente. Sin embargo, la virtud mayor de la estadística es que nos ayuda a lidiar con la incerteza y variabilidad propia del mundo natural (Sokal & Rohlf 1969). Así, más que considerar a la inferencia estadística como un proceso de extraer conclusiones, la debemos ver como una manera de aceptar y medir la incerteza y variabilidad contenida en nuestra información (Mallows 1998, Gelman 2016, 2019, Amrhein *et al.* 2017, Tong 2019).

20. Siempre reconoce explícitamente en tu manuscrito las limitaciones de tus diseños experimentales y de tus análisis estadísticos. Cuando estés consciente de tus errores estadísticos reconócelos y corrígelos rápidamente (Gelman 2018, 2020). Controla la especula-

ción. No concluyas o infieras algo más allá de lo que revelan tus resultados. Generalmente, las extrapolaciones son riesgosas en un contexto estadístico (Buckland 1980). La honestidad y transparencia es fundamental para la integridad de la ciencia (Gelman 2017, 2018, Parker *et al.* 2018). Los autores que informan adecuadamente son más creíbles. A menudo nuestros sesgos y nuestro ego nos hacen olvidar que somos fallibles y que nuestros estudios no son perfectos (Gigerenzer 1993, Lanni 2021). Admitir nuestros errores es la esencia del progreso en ciencia (Kareiva & Marvier 2018). Vuilleumier (2004) nos hizo darnos cuenta de que aprender de los errores permite tremendos avances teóricos en ornitología.

21. Mantén en tu mente los siguientes mensajes. La estadística no es una caja de herramientas para propósitos de investigación. El verdadero sentido de la estadística es aprender de la información y medir, controlar y comunicar la incerteza y variabilidad contenida en esa información (Gelman 2018, Wild *et al.* 2018). Los procedimientos estadísticos, utilizados con sencillez y modestia, nos protegen de los hallazgos falsos (Cox 2001). El uso sabio y honesto de la estadística contribuye a comprender el mundo que nos rodea y a acercarnos a la verdad contenida en él (Salkind 2017, Sedgwick 2022). La estadística nos puede ayudar a ver el mundo claramente si estamos dispuestos a mirar (Tarran 2020).

Nuestro compromiso como editores

El uso sensato de la estadística contribuye a revelar los vínculos que hay entre nuestros hallazgos y las teorías en la que enmarcamos nuestros estudios. Los resultados de los análisis estadísticos nos ayudan a evaluar nuestras hipótesis y teorías, descubrir patrones y tendencias inesperados, y a proporcionar el ímpetu para reformular las teorías con la que trabajamos (Krebs 2000, León-Guerrero & Frankfort-Nachmias 2018). Sin embargo, un procedimiento estadístico es solo una pequeña parte del proceso de investigación y ni el procedimiento más riguroso nos revelará toda la verdad que perseguimos.

Por otra parte, toda investigación científica es vulnerable al mal uso de la estadística y a los sesgos del investigador (Gelman 2018, Kareiva & Marvier 2018). El mal uso y abuso de la estadística (muchas veces involuntario) combinado con diseños experimentales deficientes ha resultado en muchos hallazgos falsos o al menos cuestionables (Ioannidis 2005, Ioannidis *et al.* 2014). Esto puede retrasar considerablemente el progreso en ornitología (Martínez-Abraín & Oro 2005). Tu deber como autor es ajustar siempre de manera apropiada y realista

tus análisis estadísticos a tu diseño experimental y razonamiento científico (Goodman 2016). Al igual que Buckland (1980), esperamos que nuestras sugerencias alienten a los ornitólogos a utilizar hábil y sensatamente la estadística y reorienten a aquellos que la usan inadecuadamente.

Los editores y revisores de revistas científicas somos los árbitros de la práctica científica (Cherry 1998, Johnson 1999, Parker *et al.* 2018, Sedgwick *et al.* 2022). Por lo tanto, no solo necesitamos comprender cómo los procedimientos estadísticos funcionan sino también transmitir su uso apropiado a los autores (Johnson 1999, Underhill 1999). Esto último incluye advertir a los autores que enmarquen sus conclusiones dentro de las limitaciones de su estudio. Si los autores presentan análisis estadísticos bien engranados con un diseño experimental sólido no tendremos mucho que decir. Sin embargo, si los autores presentan análisis estadísticos deficientes o inadecuados nuestra responsabilidad ética y profesional es sugerirles las opciones más sensatas (Parker *et al.* 2018). Esto último podría incluir incluso descartar cualquier análisis estadístico inferencial. Ante un análisis estadístico deficiente, los editores y revisores tenemos la responsabilidad de sugerir cambios y orientar apropiadamente tales análisis para garantizar la credibilidad de los hallazgos y cuidar el prestigio de los autores (Ioannidis 2014). Aunque nuestra labor es “juzgar” el mérito científico de cada manuscrito, una de nuestras premisas es no rechazar un manuscrito debido a errores analíticos que son corregibles. Como editores de la Revista Chilena de Ornitología siempre estaremos disponibles para asistir estadísticamente a los autores que los requieran.

GLOSARIO...DE ALGUNOS TÉRMINOS ESTADÍSTICAMENTE SIGNIFICATIVOS!

Alfa (α), nivel α , valor de α , “nivel de significación”: es la probabilidad de cometer un error de tipo I (ver abajo). Un investigador debe adoptar o decidir qué α usará previamente al diseño de experimentos. Podemos interpretar α como un valor numérico fijado “nominalmente” en base a la experiencia y que esperamos cometer como error. Habitualmente, los ornitólogos adoptan un $\alpha = 5\%$.

Beta (β): es la probabilidad de cometer un error de tipo II. Al igual que α , un investigador debe decidir el valor de beta en la etapa de diseño del estudio. Valores conservadores de β varían entre 10-20%.

Diseño experimental: cualquier tipo de experimento previamente planificado, ya sea un experimento bien controlado (típicamente un experimento de laboratorio), un experimento manipulativo en el campo (*e.g.*, condición natural previa vs condición intervenida posterior) o un experimento observacional o mensurativo (*i.e.*, comparación de dos o más situaciones naturales sin la intervención del investigador) (James & McCulloch 1985, Scheiner & Gurevitch 2001, Gotelli & Ellison 2004).

Error de tipo I: cometemos un error de tipo I cuando rechazamos la hipótesis de nulidad, siendo esta verdadera.

Error de tipo II: cometemos un error de tipo II cuando no rechazamos la hipótesis de nulidad siendo esta falsa.

Evidencia: cualquier hecho o pieza de información que apoya lo que creemos, postulamos o hipotetizamos (Zalta & Nodelman 2021)

Hipótesis: proposición general que sugiere explicación para un fenómeno observado (Krebs 2000). Podemos verificar una hipótesis solo de manera indirecta, examinando sus predicciones (Farji-Brener 2004).

Hipótesis alternativa (H_1): una hipótesis de diferencia o asociación entre tratamientos o condiciones experimentales, la cual indica que hay un efecto. Es la hipótesis que el investigador postula como válida en base a los patrones observados en sus datos, los que difieren con lo que postula la hipótesis de nulidad.

Hipótesis nula o de nulidad (H_0): una hipótesis de no diferencia o no asociación entre tratamientos o condiciones experimentales. Bajo la prueba de significación, esta hipótesis es admisible hasta que no haya evidencia de lo contrario y el valor establecido del parámetro sea válido. Un investigador se apoya en una hipótesis de nulidad para explicar los patrones observados en sus datos de una manera simple, atribuyendo al azar o a errores de medición cualquier variación en la información (Gotelli & Ellison 2004).

H_0 y H_1 son esencialmente hipótesis estadísticas; es decir, son conjeturas restringidas y operativas sobre el valor de cierto “parámetro” (poblacional) y representan resultados esperados bajo diferentes escenarios biológicos (Farji-Brener 2004).

Inferencia estadística: proceso de extraer conclusiones acerca de la población usando la información proveniente de una muestra o un conjunto de muestras de esa misma población. La extracción de la información proveniente de una muestra que permite la inferencia estadística es posible mediante métodos matemáticos.

Intervalo de confianza: el rango de valores alrededor del valor de un estimador estadístico (*e.g.*, estimación de media) obtenido de una muestra y que teóricamente incluirá con una alta probabilidad el valor del parámetro de la población. Usualmente, los ornitólogos estiman intervalos de confianza al 95% (Fowler & Cohen 1994). Es decir, un intervalo de confianza que incluya, por ejemplo, la media poblacional en 95 ensayos de 100. La ventaja de los intervalos de confianza es que nos permiten una evaluación cuantitativa acerca de la magnitud del efecto y de su precisión.

Paradigma: una visión del mundo; un enfoque amplio de los problemas abordados en un campo de la ciencia (Krebs 2000).

Parámetro estadístico: del griego *parametron* que significa más allá de la medición. En estadística, un parámetro es el valor de una función matemática que es capaz de resumir la información relacionada a una variable en estudio (*e.g.*, media, varianza, desviación estándar). Ya que desconocemos el valor de los parámetros de la población, calculamos ese valor mediante las mediciones obtenidas en una muestra o varias muestras de esa población mediante una función matemática. Al valor de dicha medida le llamamos “estimador”, y al cálculo específico relacionado solo a una muestra le

llamamos “estimación”. Así, no debemos confundir estimación con parámetro; un parámetro estadístico siempre estará “más allá de nuestra medición”.

Poder estadístico: es una medición de la capacidad de un experimento para detectar un efecto cuando ese efecto es real (Button *et al.* 2013, Halsey *et al.* 2015). Matemáticamente, el poder estadístico es el complemento de beta: $1 - \beta$. Así, mientras mayor es el poder estadístico, menor es la probabilidad de cometer un error de tipo II, y mayor es la posibilidad de detectar un efecto real. Aunque el nivel del poder estadístico depende de varios factores, el tamaño de la muestra es determinante. Si un estudio experimental está basado en un tamaño pequeño de muestra, entonces tendrá un bajo poder estadístico. Dependiendo del tipo de estudio, un investigador podría agrandar la magnitud del efecto cuando usa muestras demasiado pequeñas. Por esta razón, cuando uno espera que los tamaños del efecto sean débiles o moderados es recomendable evitar inferencias basadas en el umbral del valor de p cuando el tamaño de la muestra es pequeño (*e.g.*, < 10).

Predicción: las predicciones son resultados esperados bajo el supuesto de que nuestra hipótesis de investigación es verdadera. Es decir, la hipótesis estadísticas de nulidad y alternativa expresan exactamente el resultado esperado si nuestra hipótesis biológica es correcta (Farji-Brener 2004).

Significación estadística: existen muchas formas de pensar y abordar el concepto de significación estadística. Usualmente, llamamos significación estadística al valor α . Esto es relevante cuando realizamos el contraste de hipótesis, ya que el rechazo o no de la hipótesis de nulidad depende de si el valor de p queda bajo o sobre α . Damos la definición más conservadora basada en el paradigma Fisheriano y que considera el razonamiento estadístico más allá del contraste analítico:

Alcancamos la significación estadística en una muestra en estudio cuando los resultados son observables en la población sobre la cual realizamos el muestreo estadístico (probabilístico) y no se deben al azar, sino que estos responden a patrones de evidencia suficientemente consistentes como para ser observables en las muestras provenientes de dicha población.

Debido a la imposibilidad de estudiar la población, intentamos observar dicho patrón en la muestra (habitualmente una única muestra) e inferimos lo que ocurre en la población por medio de métodos estadísticos de contraste basados en pruebas de significación estadística. El propósito de la prueba de significación estadística es descartar el azar como la explicación a lo observado. Por esta razón, las muestras son aleatorias. Realizamos un proceso de aleatorización con el fin de evitar desviaciones sistemáticas inducidas por el experimentador al momento de realizar la selección de las unidades en estudio.

Tamaño del efecto: la magnitud de un efecto observado después de un tratamiento experimental. Dependiendo de la pregunta, un investigador mide el tamaño del efecto cuantificando la diferencia entre dos tratamientos o condiciones, o midiendo la fuerza de la asociación entre dos variables o procesos (Button *et al.* 2013). A menudo, los investigadores usan categorías estandarizadas del tamaño del efecto (*e.g.*, débil, moderado, fuerte; ver Cohen 1988). El efecto medido en una muestra es una estimación del tamaño del efecto “verdadero” en la población. Usualmente, interpretamos el valor de p suponiendo que el tamaño verdadero del efecto es cero

(Halsey *et al.* 2015).

Valor de p o valor- p : conceptualmente, el valor de p es la probabilidad que un estimador estadístico de los datos (*e.g.*, coeficiente de correlación) sea tan o más extremo que su valor observado (estimación) dada la hipótesis nula (Wasserstein & Lazar 2016). Dicho de otra manera, el valor de p es la probabilidad de cometer un error de tipo I si rechazamos la hipótesis de nulidad basado en el análisis de los datos. Después de comparar el valor de p con α , adoptamos la regla de que si $p < \alpha$ rechazamos H_0 (*i.e.*, retenemos H_1) y si $p > \alpha$ admitimos H_0 (*i.e.*, rechazamos H_1). La igualdad ($=$) de p y α resulta controversial y lo abordaremos posteriormente en la sección de alfabetización estadística.

Verdad: es un concepto profundamente filosófico y su definición (si es que es definible) está sujeta a diversas ideologías (Zalta & Nodelman 2021). En términos científicos podemos entender “verdad” como el conjunto de propiedades invariables o innegables de la naturaleza (*e.g.*, organismos, patrones, procesos, mecanismos) (Cohen 1980). Al menos esta es la verdad que nos interesa como ornitólogos. Muchas de las propiedades de la naturaleza son invisibles o intangibles para nosotros y buscaremos evidencia que nos acerquen a ellas. Para esto, postulamos hipótesis, diseñamos experimentos, analizamos nuestros resultados y aplicamos procedimientos inferenciales respecto a ello. Si todo es bien hecho, nuestras inferencias reflejarán algo de la verdad que buscamos.

AGRADECIMIENTOS.- Agradecemos a la Dra. Karla Yohannessen, académica de la escuela de salud pública de la Universidad de Chile, por revisar una versión previa de esta editorial. Sus comentarios y sugerencias nos permitieron refinar la redacción del texto.

LITERATURA CITADA

- AMRHEIN, A., S. GREENLAND & B. McSHANE. 2019. Retire statistical significance. *Nature* 567: 305-307.
- AMRHEIN, V., F. KORNER-NIEVERGELT & T. ROTH. 2017. The earth is flat ($p > 0.05$): significance thresholds and the crisis of unreplicable research. *PeerJ* 5: e3544. <https://peerj.com/articles/3544>
- ANDERSON, D.R. & K.R. BURNHAM. 2002. Avoiding pitfalls when using information-theoretic methods. *Journal of Wildlife Management* 66: 912-918.
- ANDERSON, D.R., K.P. BURNHAM & W.L. THOMPSON. 2000. Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management* 64: 912-923.
- ANDERSON, D.R., W.A. LINK, D.H. JOHNSON & K.P. BURNHAM. 2001. Suggestions for presenting the results of data analyses. *Journal of Wildlife Management* 65: 373-378.
- BLAIR, R.C. & R. TAYLOR. 2008. *Bioestadística*. Pearson Educación. Naucalpan de Juárez, México. 552 pp.
- BRENNAN, L. 2012. Editorial guidance and wildlife science: the roles of Wildlife Society Bulletin authors, associate editors, and reviewers. *Wildlife Society Bulletin*

- 36: 392-398.
- BUCKLAND, S.T. 1982. Statistics in ornithology. *Ibis* 124: 61-66.
- BURNHAM, K.P. & D.R. ANDERSON. 2001. Kullback–Leibler information as a basis for strong inference in ecological studies. *Wildlife Research* 28: 111-119.
- BUTTON, K.S., J.P.A. IOANNIDIS, C. MOKRYSZ, B.A. NOSEK, J. FLINT, E.S.J. ROBINSON & M.R. MUNAFO. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 14: 365-376.
- CARRAWAY, L.N. 2009. Ethics for and responsibilities of authors, reviewers and editors in science. *American Midland Naturalists* 161:146-164.
- CHAMBERLIN. T.C. 1890. The method of multiple working hypotheses. *Science* 148: 754-759.
- CHERRY, S. 1998. Statistical tests in publications of the Wildlife Society. *Wildlife Society Bulletin* 26: 947-953.
- CHIA, K-S. 1997. “Significant-itis” – an obsession with the P-value. *Scandinavian Journal of Work Environmental Health* 23: 152-154.
- COHEN, J. 1980. What has science to do with truth? *Synthese* 45: 489-510.
- COHEN, J. 1988. Statistical power analysis for the behavioral sciences. Segunda edición. Lawrence Erlbaum Associates, Nueva York, EE.UU. 567 pp.
- COLLING, L.J. & D. SZÜCS. 2021. Statistical inference and the replication crisis. *Review of Philosophy and Psychology* 12: 121-147.
- COMROY, G. 2019a. The 7 deadly sins of research. *Nature News*. <https://www.nature.com/nature-index/news-blog/the-seven-deadly-sins-of-research>
- COMROY, G. 2019b. The 10 most common mistakes with statistics, and how to avoid them. *Nature News*. <https://www.nature.com/nature-index/news-blog/ten-most-common-mistakes-with-statistics-and-how-to-avoid-them>
- COX, D.R. 2001. Another comment on the role of statistical methods. *Physical Therapy* 81: 1469.
- CUMMING, G. 2012. *Understanding the new statistics - effect sizes, confidence intervals, and meta-analysis*. Routledge Taylor & Francis Group, Nueva York, EE.UU. 519 pp.
- EGLI, G. 1996. Biomorfología de algunas aves de Chile central. *Boletín Chileno de Ornitología* 3: 2-9.
- EKLUND, A., T.E. NICHOLS & H. KNUTSSON. 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of National Academy of Sciences (PNAS)* 113: 7900-7905.
- ELLISON, A.M. 2001. Exploratory data analysis and graphic display. Pp. 37-62, en Scheiner, S.M. & J. Gurevitch (eds.) *Design and analysis of ecological experiments*. Oxford University Press, Nueva York, EE.UU.
- ELLISON, A.M. 2004. Bayesian inference in ecology. *Ecology Letters* 7: 509-520.
- FEINSINGER, P. 2001. *Designing field studies for biodiversity conservation*. The Nature Conservancy. Island Press, Washington, EE.UU. 212 pp.
- FARJI-BRENER, A. 2004. ¿Son hipótesis las hipótesis estadísticas? *Ecología Austral* 14: 201-203.
- FIELD, A. 2016. *An adventure in statistics - the reality enigma*. SAGE Publications, Londres, Reino Unido. 746 pp.
- FISHER, R. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London (Series A)* 222: 594-604.
- FISHER, R. 1935. *The design of experiments*. Oliver and Boyd, Edinburgh, Reino Unido. 257 pp.
- FISHER, R. 1956. Statistical methods and scientific induction. *Journal of the Royal Statistical Society (Serie B)* 17: 69-78.
- FOWLER, J. & L. COHEN. 1994. *Statistic for ornithologists*. Segunda edición. British Trust for Ornithology, Londres, Reino Unido. 150 pp.
- GELMAN, A. 2016. The problems with P-values are not just with P-values. Supplemental material to the ASA statement on P-values and statistical significance. *American Statistician* 70: 129-133.
- GELMAN, A. 2017. Honesty and transparency are not enough. *Chance* 30: 37-39.
- GELMAN, A. 2018. Ethics in statistical practice and communication: five recommendations. *Significance* 15: 40-43.
- GELMAN, A. 2019. The principles of uncertainty. *Nature* 569: 628-629.
- GELMAN, A. 2020. Evidence vs. truth. <http://www.stat.columbia.edu/~gelman/research/published/ethics22.pdf>
- GELMAN, A. & E. LOKEN. 2014. The statistical crisis in science. *American Scientists* 102: 460-465.
- GELMAN, A., J. HILL, & M. YAJIMA. 2012. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* 5: 189-211.
- GERRARD, J. 1985. Statistical analysis in ornithology. *Ring-ing & Migration* 6: 60.
- GIGERENZER, G. 1993. The superego, the ego, and id in statistical reasoning. Pp. 311-339, en Keren, G. & C. Lewis (eds.) *A handbook for data analysis in the behavioral sciences: methodological issues*. Lawrence Erlbaum Associates, Nueva Jersey, EE.UU.
- GOODMAN, S.N. 2016. Aligning statistical and scientific reasoning - misunderstanding and misuse of statistical significance impede science. *Science* 352: 1180-1181.
- GOTELLI, N.J. & A.M. ELLISON. 2004. *A primer of ecological statistics*. Sinauer Associates Inc., Sunderland, Massachusetts, EE.UU. 510 pp.
- GREENLAND, S. & C. POOLE. 2013. Living with statistics in observational research. *Epidemiology* 24: 73-78.
- GUSTAVII, B. 2008. *How to write and illustrate scientific papers*. Segunda edición. Cambridge University Press. Cambridge, Reino Unido. 168 pp.

- HALSEY, L.G., D. CURRAN-EVERETT, S.L. VOWLER & G.B. DRUMMOND. 2015. The fickle *P* value generates irreproducible results. *Nature Methods* 12:179-185.
- HAMPEL, F. 1997. *Is statistic too difficult?* Research Report N° 81. Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), Zürich, Suiza. 18 pp.
- HUFF, D. 1993. *How to lie with statistics*. W.W. Norton & Company Inc., Nueva York, EE.UU.
- IN, J. & S. LEE. 2017. Statistical data presentation. *Korean Journal of Anesthesiology* 70: 267-287.
- IOANNIDIS, J.P.A. 2005. Why most published research findings are false. *Public Library of Science (PLOS) Medicine* 2: e124.
- IOANNIDIS, J.P.A. 2014. How to make more published research true. *Public Library of Science (PLOS) Medicine* 10: e1001747.
- IOANNIDIS, J.P.A., S. GREENLAND, M.A. HLATKY, M.J. KHOURY, M.R. MACLEOD, D. MOHER, K.F. SCHULZ & R. TIBSHIRANI. 2014. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 383: 166-175.
- JAMES, F.C. & C.E. MCCULLOCH. 1985. Data analysis and the design of experiments in ornithology. Pp. 1-63, en Johnston, R.F. (ed.) *Current Ornithology*. Plenum Press, Nueva York, EE.UU.
- JAMES, F.C. & C.E. MCCULLOCH. 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology and Systematics* 21:129-166.
- JOHNSON, D.H. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* 63: 763-772.
- KAMIL, A.C. 1988. Experimental design in ornithology. Pp. 313-346, en Johnston, R.F. (ed.) *Current Ornithology*. Plenum Press, Nueva York, EE.UU.
- KAREIVA, P. & M. MARVIER. 2018. Uncomfortable questions and inconvenient data in conservation science. Pp. 3-9, en Kareiva, P., M. Marvier & B. Silliman (eds.) *Effective conservation science: data not dogma*. Oxford University Press, Oxford, Reino Unido. 190 pp.
- KREBS, C.J. 1989. *Ecological methodology*. Harper Collins Publisher. Nueva York, EE.UU. 654 pp.
- KREBS, C.J. 2000. Hypothesis testing in ecology. Pp. 5-14, en Boitani, L. & T. Fuller (eds.) *Research techniques in animal ecology: controversies and consequences*. Columbia University Press, Nueva York, EE.UU.
- LANG, J.M., K.J. ROTHMAN & C.L. CANN. 1998. That confounded *P*-value. *Epidemiology* 9: 7-8.
- LANNI, J. 2021. A lesson of failure. *Science* 374: 1642.
- LECOUTRE, B., J. POITEVINEAU & M-P. LECOUTRE. 2005. A reason why not to ban null hypothesis significance tests. *Revue MODULAD* 33: 249-253.
- LEÓN-GUERRERO, A. & C. FRANKFORT-NACHMIAS. 2018. *Essentials of social statistics for a diverse society*. Tercera edición. SAGE Publications, Thousand Oaks, California, EE.UU. 547 pp.
- LITTLEWOOD, B. & L. STRIGIN. 1992. The risks of software. *Scientific American* 267: 62-75.
- MALAY, D.S. 2016. Editorial: clinical significance, clarified. *Journal of Foot & Ankle Surgery* 55: 903-905.
- MALLOWS, C.L. 1998. The zeroth problem. *American Statistician* 52: 1-9.
- MARTÍNEZ-ABRAÍN, A. & D. ORO. 2005. Can ornithology advance as a science relying on significance testing? A literature review in search of a consensus. *Ardeola* 52: 377-387.
- MCELREATH, R. 2016. *Statistical rethinking - a Bayesian course with examples in R and Stan*. Chapman & Hall Book. Boca Raton, Florida, EE.UU. 469 pp.
- MAKIN, T.R. & J.J. ORBAN DE XIVRY. 2019. Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *eLife* 8: e48175.
- NELDER, J.A. 1999. From statistics to statistical science. *Statistician* 48: 257-269.
- NORTH, P.M. & J.T. BYRON. 1985. The general context of statistics in ornithology. Pp. 1-9, en Morgan, B.J.T. & P.M. North (eds.) *Statistics in ornithology*. Springer-Verlag, Berlin, Alemania.
- NORTH, P.M. 1994. Ornithological statistics. Pp. 463-506, en Patil, G.P. & C.R. Rao (eds.) *Handbook of statistics. Vol. 12. Environmental Statistics*. Elsevier Science Publisher. Cambridge, Massachusetts, EE.UU.
- PARKER, T.H., S.C. GRIFFITH, J.L. BRONSTEIN, F. FIDLER, S. FOSTER, H. FRASER, W. FORSTMEIER, J. GUREVITCH, J. KORICHEVA, R. SEPPELT, M.W. TINGLEY & S. NAKAGAWA. 2018. Empowering peer reviewers with a checklist to improve transparency. *Nature Ecology & Evolution* 2: 929-935.
- POTISH, R., J. BOEN & L. SEYMOUR. 1980. The distinction between statistical significance and clinical importance. *International Journal of Radiation Oncology, Biology and Physics* 6: 941-944.
- QUINN, G.P. & K.J. KEOUGH. 2002. *Experimental design and data analysis for biologists*. Cambridge University Press, Londres, Reino Unido. 553 pp.
- SALKIND, N.J. 2017. *Statistics for people who (think they) hate statistics: using Microsoft Excel 2016*. Cuarta edición. Sage Publications. California, EE.UU. 512 pp.
- SCHEINER, S.M. & J. GUREVITCH (eds.). 2001. *Design and analysis of ecological experiments*. Oxford University Press, Nueva York, EE.UU. 415 pp.
- SCHEINER, S.M. 2001. Theories, hypotheses, and statistics. Pp. 3-13, en Scheiner, S.M. & J. Gurevitch (eds.) *Design and analysis of ecological experiments*. Oxford University Press, Nueva York, EE.UU.
- SEDGWICK, P. 2022. Trials and tribulations of teaching null hypothesis significance testing in the health sciences. *Chance* 35: 33-41.
- SEDGWICK, P., A. HAMMER, U.S. KESMODEL & L.H. PEDERS-

- EN. 2022. Current controversies: null hypothesis significance testing. *Acta Obstetricia et Gynecologica Scandinavica* 101: 624-627.
- SIEGEL, S. & N.J. CASTELLAN. 1995. *Estadística no paramétrica aplicada a las ciencias de la conducta*. Cuarta edición. Editorial Trillas, México D.F., México. 437 pp.
- SIMMONS, J.P., L.D. NELSON & U. SIMONSOHN. 2013. *Life after P-hacking*. Meeting of the Society for Personality and Social Psychology. New Orleans, Louisiana, EE.UU. 17-19 January 2013. <http://dx.doi.org/10.2139/ssrn.2205186>
- SOKAL, R.R. & F.J. ROHLF. 1969. *Biometry*. Freeman & Company, San Francisco, EE.UU. 776 pp.
- STEIDL, R.J. & L. THOMAS. 2001. Power analysis and experimental design. Pp. 14-36, *en* Scheiner, S.M & J. Gurevitch (eds.) *Design and analysis of ecological experiments*. Oxford University Press, Nueva York, EE.UU.
- TARRAN, B. 2020. "Statistics help us see the world clearly if we're willing to look". *Significance* 17: 42-43.
- TONG, C. 2019. Statistical inference enables bad science; statistical thinking enables good science. *American Statistician* 73: 246-261.
- UNDERHILL, L.G. 1985. Avian demography: statistics and ornithology. *Ostrich* 70: 61-70.
- VAN DE SCHOOT, R., S. DEPAOLI, R. KING, B. KRAMER, K. MÄRTENS, M.G. TADESSE, M. VANNUCCI, A. GELMAN, D. VEEN, J. WILLEMSSEN & C. YAU. 2021. Bayesian statistics and modelling. *Nature Reviews* 1: 1. <https://doi.org/10.1038/s43586-020-00001-2>
- VAN EMDEN, H.F. 2008. *Statistics for terrified biologists*. Blackwell Publishing, Oxford, Reino Unido. 343 pp.
- VUILLEUMIER, F. 2004. A critique of Neotropical ornithology: is research on neotropical birds scientific? *Ornitología Neotropical* 15(Suppl.): 41-60.
- WASSERSTEIN, R.L. & N.A. LAZAR. 2016. The ASA's statement on p-values: context, process, and purpose. *American Statistician* 70: 129-133.
- WHITE, J.W., A. RASSWEILER, J.F. SAMHOURI, A.C. STIER & C. WHITE. 2014. Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos* 123: 385-388.
- WILD, C.J., J.M. UTTS & N.J. HORTON. 2018. What is statistics? Pp. 5-36, *en* Ben-Zvi, D., K. Makar & J. Garfield (eds.) *International handbook of research in statistics education*. Springer International Publishing, Suiza. https://doi.org/10.1007/978-3-319-66195-7_1.
- YOCOZ, N.G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* 72: 106-111.
- ZALTA, E.N. & U. NODELMAN. 2021. *Stanford encyclopedia of philosophy*. Department of Philosophy, Stanford University, California, EE.UU. <https://plato.stanford.edu/>
- ZAR, J.H. 1999. *Biostatistical analysis*. Cuarta edición. Prentice Hall, Upper Saddle River, Nueva Jersey, EE.UU. 663 pp.
- ZIEFFLER, A. J. GARFIELD & E. FRY. 2018. What is statistics education? Pp. 37-70, *en* Ben-Zvi, D., K. Makar & J. Garfield (eds.) *International handbook of research in statistics education*. Springer International Publishing, Suiza. https://doi.org/10.1007/978-3-319-66195-7_2.