

ORIGINAL RESEARCH

Aplicación de la ley de Benford a la pesquería argentina de corvina rubia (*Micropogonias furnieri*)

SEBASTIÁN GARCÍA*, JULIETA S. RODRÍGUEZ y BRUNO V. MENNA

¹Instituto Nacional de Investigación y Desarrollo Pesquero (INIDEP), Paseo Victoria Ocampo N° 1, Escollera Norte, B7602HSA - Mar del Plata, Argentina. ORCID *Sebastián García*  <https://orcid.org/0009-0004-6390-5090>



RESUMEN. En el ámbito de los datos estadísticos, tanto en las ciencias naturales como sociales, se ha observado que la distribución de los primeros, segundos y dos primeros dígitos en datos reales sigue frecuentemente un patrón conocido como la “ley de Benford”. Esta ley, ha sido utilizada recientemente como una herramienta para identificar anomalías en distintas bases de datos, sugiriendo en algunos casos la posibilidad de fraude. Se observó que los dígitos de datos “genuinos” tienden a seguir la ley, mientras que los dígitos de datos manipulados no lo hacen. En este trabajo, exploramos su aplicabilidad más allá del ámbito financiero, investigando si pueden detectar irregularidades en datos científicos, específicamente en la estadística oficial de captura de la pesquería argentina de corvina rubia (*Micropogonias furnieri*). Para tal fin, comparamos la frecuencia del primer, segundo y el primer par de dígitos de la captura con la distribución esperada, utilizando la desviación media absoluta (MAD). Implementamos una metodología basada en simulaciones de Monte Carlo y el test de Kolmogorov-Smirnov para calcular los valores críticos de la prueba de conformidad MAD, abordando la naturaleza única de los datos y la variabilidad en el tamaño de la muestra. El análisis realizado, sugirió la existencia de anomalías que podrían indicar patrones inusuales que merecen una investigación más detallada. En el ámbito de la evaluación, manejo/administración y conservación de los recursos pesqueros, la confiabilidad de los datos de captura es esencial. El uso de la ley de Benford, podría optimizar la selección de la información utilizada para elaborar indicadores y reducir la incertidumbre en la estimación del estado poblacional de los recursos.

Palabras clave: Desviación media absoluta (MAD), manejo de pesquerías, recursos pesqueros.



*Correspondence:
sgarcia@inidep.edu.ar

Received: 11 March 2024
Accepted: 24 May 2024

ISSN 2683-7595 (print)
ISSN 2683-7951 (online)

<https://ojs.inidep.edu.ar>

Journal of the Instituto Nacional de
Investigación y Desarrollo Pesquero
(INIDEP)



This work is licensed under a Creative
Commons Attribution-
NonCommercial-ShareAlike 4.0
International License

Application of Benford’s Law to the Argentine whitemouth croaker (*Micropogonias furnieri*) fishery

ABSTRACT. In the field of statistical data, both in the natural and social sciences, it has been observed that the distribution of the first, second and first two digits in real data frequently follows a pattern known as ‘Benford’s law’. This law has recently been used as a tool to identify anomalies in different databases, suggesting in some cases the possibility of fraud. It was observed that ‘genuine’ data digits tend to follow the law, while manipulated data digits do not. In this work, we explore their applicability beyond the financial sphere, investigating whether they can detect irregularities in scientific data, specifically in the official catch statistics of the Argentine whitemouth croaker (*Micropogonias furnieri*) fishery. To this end, we compare the frequency of the first, second and first pair of digits of the capture with the expected distribution using the mean absolute deviation (MAD). We implemented a methodology based on Monte Carlo simulations and the Kolmogorov-Smirnov test to calculate critical values of the MAD conformance test, addressing the unique nature of data and the variability in sample size. The conducted analysis suggested the presence of anomalies that could denote unusual patterns warranting further detailed investigation. In the field of assessment,

management/administration, and conservation of fishing resources, the reliability of catch data is essential. The use of Benford's law could optimize the selection of information used to develop indicators and reduce uncertainty in estimating the population status of resources.

Key words: Mean absolute deviation (MAD), fisheries management, fisheries resources.

INTRODUCCIÓN

La pesca comercial en las costas argentinas se centra en un grupo de especies conocido como variado costero (Carozza et al. 2001). Dentro de este grupo, se encuentra la corvina rubia (*Microponias furnieri* Desmarest, 1823), una especie demersal de tamaño mediano cuya zafra se realiza principalmente en aguas de la Bahía Samborombón ($35^{\circ} 51' S-57^{\circ} 7' W$) entre junio y octubre de cada año. Los desembarques de esta especie representan cerca de 31% del total del conjunto íctico demersal costero (Carozza et al. 2019). La modalidad de pesca es variada, incluyendo líneas de mano, redes de enmalle y redes de arrastre de fondo con portones o mediante el método "a la pareja" que implica el uso coordinado de dos embarcaciones. La flota que explota este recurso se concentra principalmente en los puertos de Mar del Plata ($38^{\circ} 0' S, 57^{\circ} 33' W$) y General Lavalle ($36^{\circ} 25' S, 56^{\circ} 57' W$). Se compone de más de 150 embarcaciones de diversos tamaños, desde pequeños barcos artesanales de fibra de vidrio con esloras menores a ocho metros hasta embarcaciones de gran porte con esloras que superan los 38 m y potencias de hasta 800 HP (Carozza 2012; García 2023). El segmento más significativo, en términos de volúmenes desembarcados, corresponde al denominado estrato IC (18 a 25 m de eslora) que representó más de 53% del total anual desembarcado de la especie en el período 2002-2018 (Carozza et al. 2019).

La gestión sostenible de este tipo de recursos es fundamental en el contexto de la seguridad alimentaria y depende en gran medida de un sistema de toma de decisiones basado en información confiable. Sin embargo, la confiabilidad de las estadísticas oficiales de pesca ha sido objeto de un amplio

debate en la literatura científica (Teschmichael y Pauly 2011; Pauly et al. 2014). Índices relativos de abundancia, ubicación de las áreas frecuentes de captura, estacionalidad de los desembarques o la fauna acompañante en cada viaje, son algunos indicadores de estado que provienen de la información declarada en los denominados "partes de pesca". Sin embargo, y a pesar del carácter de declaración jurada que revisten estos documentos, se han detectado innumerables inconsistencias, errores y omisiones en su elaboración y presentación a la autoridad de gobierno (García et al. 2018). Los actuales modelos de evaluación de abundancia y las proyecciones de biomasa se basan en gran medida en los datos proporcionados en los partes de pesca (Maunder y Punt 2004). Si estos datos contienen inconsistencias, los resultados de los modelos podrían ser sesgados o incorrectos, lo que comprometería la eficacia de las medidas de manejo sugeridas.

Esta situación ha forzado la búsqueda de alternativas al empleo de dicha información (García et al. 2018) y un cuidadoso trabajo de filtrado y selección de los datos que se utilizan, a menudo con un costoso tiempo de ejecución debido al volumen que presentan las bases de datos de este tipo. Los avances en el campo de la informática y el aumento del poder de cálculo permitieron la utilización de información a gran escala en la gestión de recursos. Consecuentemente, se acrecentaron los requerimientos de herramientas que permitan validar la información utilizada, máxime cuando de ella dependen, por ejemplo, la recomendación anual de captura biológicamente aceptable de una especie (CBA).

En el análisis de datos numéricos, tanto de origen natural como social, se ha verificado que la frecuencia con la que aparecen los primeros dígitos suele ajustarse a un patrón predecible. Este patrón es conocido como la ley de Benford, también lla-

mada ley del Primer Dígito o ley de los Números Anómalos, la cual fue descrita empíricamente por primera vez por Simón Newcomb en 1881 (Newcomb 1881) y formalizada posteriormente por Frank Benford (Benford 1938). Esta distribución característica ha cobrado relevancia como instrumento analítico para la detección de irregularidades en amplias bases de datos, insinuando en ocasiones la existencia de manipulación o fraude (Cabeza García 2019). La hipótesis central propone que los dígitos de conjuntos de datos genuinos tienden a seguir la distribución predicha por Benford, mientras que en los datos alterados o fabricados suelen desviarse. A pesar de que la distribución de frecuencias de Benford no es intuitiva debido a su naturaleza no uniforme, ha sido observada en una amplia variedad de conjuntos de datos, incluyendo áreas como finanzas, auditorías, dimensiones geográficas, fraudes electorales, reportes oficiales de casos de COVID-19 y declaraciones de impuestos, entre otros (Diekmann 2007; Mebane 2008; Silva y Filho 2021). En el ámbito de la biología pesquera, la implementación de técnicas estadísticas basadas en la ley de Benford han sido utilizadas para la verificación de los reportes de capturas de la langosta del Atlántico canadiense (*Homarus americanus*, Graham et al. 2010), de ballenas en el océano Pacífico (Cerri 2018), atunes en el Golfo de Guinea (Tzagbey et al. 2017) y, recientemente, para validar la información de captura de la pesca artesanal en el noreste de Brasil (Noletto-Filho et al. 2022). Sin embargo, la utilización de este tipo de herramientas no es ajena a controversias concernientes con la magnitud de las desviaciones observadas a la ley y el tamaño de la muestra (eventos *cry wolf* o exceso de potencia). La descomposición de la información en subconjuntos más pequeños y su relación con errores de tipo I es analizado extensamente por Barney y Schulzke (2016).

En función de lo anteriormente expuesto, resulta crucial abordar los desafíos relacionados con la calidad de la información y buscar alternativas para validar y mejorar la confiabilidad de los datos empleados en los estudios de gestión sostenible de

recursos. En este trabajo, se utilizó la metodología de detección estadística basada en los postulados de la ley de Benford para valorar los reportes de captura de corvina rubia (*M. furnieri*) declarados por la flota comercial argentina que operó bajo la modalidad “a la pareja”. Se propone un enfoque metodológico que intenta abordar la problemática asociada con la correlación entre el tamaño de muestra y la prueba de conformidad utilizada.

MATERIALES Y MÉTODOS

La ley de Benford es una herramienta utilizada actualmente en la detección de fraudes y manipulaciones en grandes bases de datos (Nigrini 2012). Esta ley se basa en que la ocurrencia de los dígitos en una cifra positiva posee una probabilidad logarítmica. Particularmente, para el primer dígito significativo distinto de cero (d_1), el segundo dígito significativo (d_2) y la cifra formada por el primer y segundo dígito (d_1-d_2), en una determinada muestra de números reales positivos la probabilidad de ocurrencia está determinada por las siguientes fórmulas:

$$p(d_1) = \log_{10} \left(1 + \frac{1}{d_1} \right) \text{ con } 1 \leq d_1 \leq 9$$

$$p(d_2) = \sum_{j=1}^9 \log_{10} \left(1 + \frac{1}{10j + (d_2 - 1)} \right) \text{ con } 1 \leq d_2 \leq 10$$

$$p(d_1-d_2) = \log_{10} \left(1 + \frac{1}{d_1-d_2} \right) \text{ con } 10 \leq d_1-d_2 \leq 99$$

De acuerdo con estas fórmulas, los números siguen un patrón regular cuando estos se generan de manera natural, siendo la probabilidad de ocurrencia esperada para “1” como primer dígito en un determinado conjunto de números $p_{(1)} = 0,301$, hasta $p_{(9)} = 0,045$ para “9”. Para el segundo dígito la probabilidad de ocurrencia de “0” es $p_{(0)} = 0,119$, hasta $p_{(9)} = 0,084$ para la ocurrencia de “9”.

Mientras que, para el primer y segundo dígito, la probabilidad de ocurrencia de “10” es $p_{(10)} = 0,041$ hasta $p_{(99)} = 0,004$ para “99”.

En Schröpfer (2010) se resumen algunos requisitos para poder evaluar la conformidad de una serie de datos respecto de la ley de Benford. Se destaca que los mismos no deben contener valores máximos preestablecidos ya que la ley se fundamenta en patrones de distribución natural de dígitos. El conjunto de datos debe tener valores positivos con una distribución unimodal y una distribución asimétrica positiva en la que la mediana no sea mayor a la mitad de la media, indicando una tendencia de los dígitos hacia valores menores (Scott y Fasli 2001). Además, los datos no deben ser pre-asignados, como por ejemplo números de cuentas bancarias (Nigrini 1999) ni surgir de procedimientos estadísticos como promedios o variaciones que proceden de otros datos (Mochty 2002).

Origen de la información

Teniendo en cuenta las restricciones mencionadas anteriormente, se recopilaron datos de captura nula de corvina rubia (en kilogramos) provenientes de las declaraciones oficiales de desembarque efectuadas por 70 buques de 18 a 25 m de eslora pertenecientes a la flota comercial argentina entre 2004 y 2021. Estos buques, utilizaron redes de arrastre de fondo bajo la modalidad “a la pareja” y operaron mayormente desde el puerto de Mar del Plata (Argentina). La duración promedio de los viajes de pesca de estos buques fue de ocho días, mientras que la potencia de sus motores varió en un rango de 320 a 652 HP.

Análisis estadístico

La distribución de frecuencias de los primeros dos dígitos significativos (d_1 y d_2) y de ambos en conjunto (d_1-d_2) de la captura declarada en cada año fue comparada gráfica y analíticamente con aquella esperada según los postulados de la ley de Benford. El ajuste de los datos observados a la dis-

tribución esperada se analizó mediante la técnica sugerida por Nigrini (2012), que utiliza la desviación media absoluta media (MAD por sus siglas en inglés):

$$MAD = \frac{\sum_{i=1}^k |obs_i - esp_i|}{k}$$

donde k es la cantidad de categorías que toma la variable, obs_i la frecuencia relativa observada y esp_i la frecuencia relativa de la distribución teórica de Benford (1938). Las categorías en cada análisis son las siguientes: 1, 2, ..., 9 para d_1 ; 0,1, ..., 9 para d_2 y 10, 11, ..., 99 para d_1-d_2 . Con el objeto de utilizar el valor de MAD como una prueba de conformidad de un conjunto de datos es necesario contar con un valor crítico. Si bien Nigrini (2012) estimó valores críticos a partir de un amplio conjunto de datos de diferentes orígenes, no existen estimaciones específicas para datos de pesquerías. Por otro lado, considerar un valor crítico independiente del tamaño de muestra puede llevar a conclusiones sesgadas (Cerqueti y Lupi 2021). En este contexto, resultó evidente la necesidad de estimar los valores críticos específicos para lo cual se implementó un procedimiento basado en simulaciones de Monte Carlo similar al utilizado por Kössler et al. (2021). Esta metodología se basa en el test de Kolmogorov-Smirnov discreto (KS), que utiliza el valor crítico ajustado a distribuciones discretas estimado por Morrow (2014) para un nivel de significancia del 95% ($\alpha = 0,05$). El procedimiento consistió en la generación de subconjuntos de los datos de tamaño muestral variable (sorteado entre 300 y 1.000) utilizando una función para generar números pseudoaleatorios con distribución de probabilidad uniforme. Para cada subconjunto se calculó el valor observado de MAD para el primer dígito (d_1), segundo dígito (d_2) y los primeros dos dígitos significativos (d_1-d_2). Se calculó el valor observado de KS y se comparó con el valor crítico ajustado por Morrow (2014). A partir de la conclusión alcanzada en el paso anterior, se asoció al valor observado de MAD la categoría aceptado o rechazado. Este

procedimiento se repitió 50.000 veces, número establecido de manera experimental para obtener estabilidad en los resultados hasta la tercera posición decimal. Una vez obtenidas las distribuciones de frecuencias finales de valores aceptados y rechazados, se determinó el valor crítico a partir del punto de corte óptimo. Esta estimación no paramétrica utilizada para distribuciones empíricas se obtuvo a partir de la curva de característica operativa del receptor (ROC) y el índice de Youden, metodología utilizada comúnmente en investigaciones del área biomédica (Schisterman 2005). Finalmente, al momento de evaluar un conjunto de datos, se determina la conformidad con la ley de Benford si se cumple la condición de que el valor observado de MAD sea menor o igual al valor crítico.

El análisis y el procedimiento de simulación se implementó en el entorno de programación R (R Core Team 2019).

RESULTADOS

Las frecuencias de los reportes anuales de captura de corvina rubia (*M. furnieri*) presentaron una distribución unimodal con asimetría positiva en la mayoría de los casos (Figura 1), por lo cual los datos cumplirían los requerimientos para la aplicación de un análisis basado en la ley de Benford (Scott y Fasli 2001; Mochty 2002).

Se estimaron los valores críticos anuales y total de la prueba MAD en cada uno de los tres enfoques utilizados para analizar la conformidad con la ley de Benford de la captura reportada por la flota comercial argentina en el período 2004-2021: test del primer dígito (d_1), segundo dígito (d_2) y el primer par de dígitos (d_1-d_2). Para d_1 , los valores críticos estimados se encontraron entre 0,0111 (2015) y 0,0241 (2007) mientras que para d_2 , las estimaciones variaron entre 0,0116 (2014) y 0,0249 (2006). En ambos casos, los rangos observados incluyen los valores de referencia observados por Nigrini (2012). En cuanto a d_1-d_2 , los valores críticos es-

timados estuvieron en el rango de 0,0026 (2013) a 0,0046 (2005), superando el valor de referencia previo (0,0022).

Los resultados obtenidos en la aplicación de los tres enfoques (d_1 , d_2 y d_1-d_2) para la captura declarada de corvina revelaron variaciones a lo largo del período considerado y diferencias en la conformidad entre los diferentes enfoques. Aunque en general los datos mostraron conformidad con la ley de Benford, se observaron desviaciones en ciertos años, lo que podría indicar la presencia de anomalías o irregularidades en las declaraciones de captura.

Análisis del primer dígito (d_1)

El análisis del primer dígito, reveló conformidad con la ley de Benford para la totalidad de los años en conjunto ($MAD_{\text{observado}} = 0,0080$; $MAD_{\text{crítico}} = 0,0108$). Sin embargo, al realizar un análisis discriminado por año, se observaron diferencias a lo largo del tiempo con un aumento progresivo en la conformidad a partir de 2009 seguido de una disminución en el período 2019 a 2021 (Figura 2; Tabla 1). El máximo valor observado de todo el período (0,026) correspondió a 2005 mientras que el mínimo (0,004) a 2015.

Análisis del segundo dígito (d_2)

El análisis del segundo dígito significativo (Figura 3; Tabla 1), mostró resultados similares al análisis del primer dígito, con excepción de 2004 y 2005. Los valores observados de MAD_{d_2} resultaron, en su mayoría, menores que los valores críticos correspondientes, lo que indicó conformidad con la ley de Benford en términos de la segunda posición decimal. La tendencia anual de $MAD_{\text{observado}}$ presentó conformidad en el período posterior a 2008 aunque 2018 y 2020 se ubicaron cerca de los límites establecidos (Figura 4). Además, se observó conformidad a la ley de Benford para el total de los datos (total), con un valor de $MAD_{\text{observado}} = 0,0082$ ($MAD_{\text{crítico}} = 0,0123$).

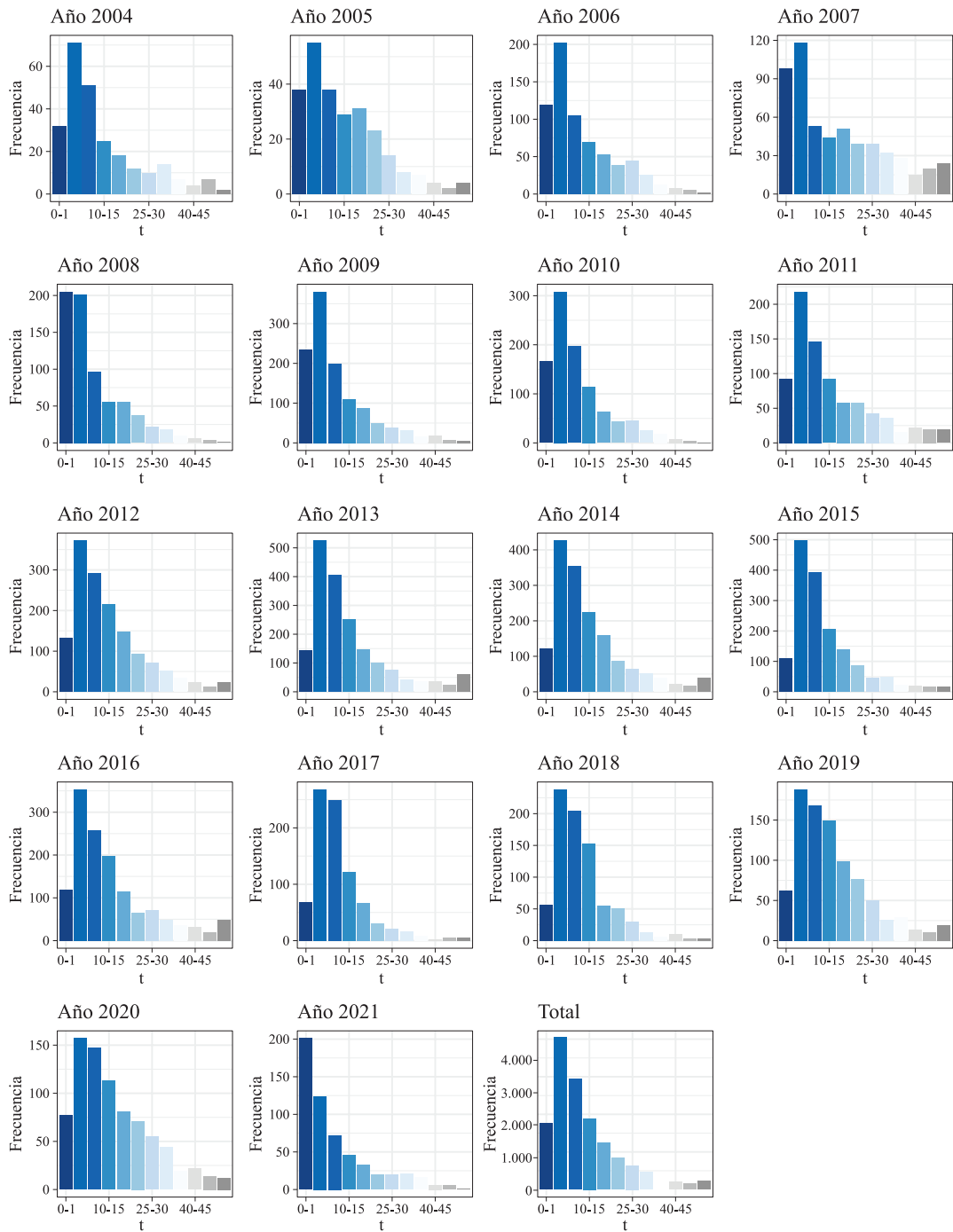


Figura 1. Histograma de frecuencias de las capturas de corvina rubia (*Micropogonias furnieri*) declaradas por la flota comercial argentina que operó “a la pareja” en el período 2004-2021.
Figure 1. Frequency histogram of Argentine whitemouth croaker (*Micropogonias furnieri*) catches reported by the Argentine commercial fleet operating in pairs from 2004 to 2021.

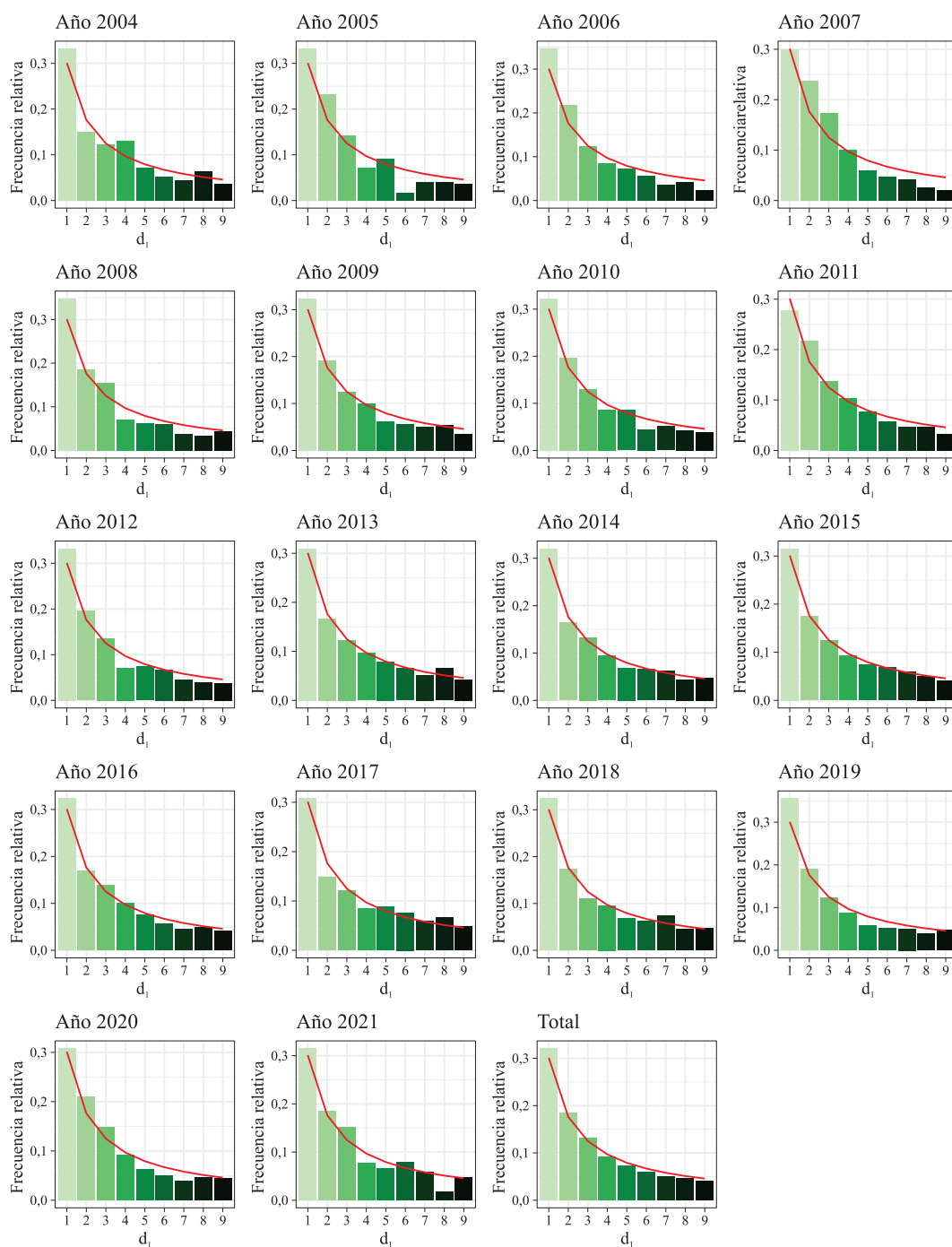


Figura 2. Distribución de frecuencia observada en el primer dígito (d_1) de la declaración oficial de captura de corvina rubia (*Micropogonias furnieri*) en el período 2004-2021 (barras) y la distribución teórica esperada de Benford (línea).

Figure 2. Observed frequency distribution of the first digit (d_1) in the official catch declaration of Argentine whitemouth croaker (*Micropogonias furnieri*) during the period 2004-2021 (bars) and the expected theoretical Benford distribution (line).

Tabla 1. Tamaño de muestra, valores observados y críticos de la prueba de conformidad MAD (desviación media absoluta) para el primer dígito (d_1), segundo dígito (d_2) y el primer par de dígitos (d_1-d_2) de la captura (kg) de corvina rubia (*Micropogonias furnieri*) reportada por la flota comercial argentina. Nivel de significancia del 95% ($p = 0,05$). En negrita se indican los valores que exceden el umbral crítico de conformidad.

Table 1. Sample size, observed and critical values of the MAD (mean absolute deviation) conformity test for the first digit (d_1), second digit (d_2), and first pair of digits (d_1-d_2) in the catch (kg) of the Argentine whitemouth croaker (*Micropogonias furnieri*) reported by the Argentine commercial fleet. Significance level: 95% ($p = 0.05$). Values exceeding the critical conformity threshold are highlighted in bold.

Año	n	MAD _{observado}			MAD _{crítico}		
		MAD _{d₁}	MAD _{d₂}	MAD _{d₁-d₂}	MAD _{d₁}	MAD _{d₂}	MAD _{d₁-d₂}
2004	253	0,0170	0,0205	0,0049	0,0185	0,0203	0,0044
2005	253	0,0260	0,0111	0,0051	0,0225	0,0142	0,0046
2006	684	0,0194	0,0261	0,0046	0,0176	0,0249	0,0040
2007	561	0,0249	0,0252	0,0049	0,0241	0,0245	0,0042
2008	713	0,0196	0,0229	0,0049	0,0180	0,0213	0,0045
2009	1.181	0,0104	0,0181	0,0032	0,0128	0,0202	0,0033
2010	993	0,0118	0,0148	0,0029	0,0131	0,0166	0,0029
2011	821	0,0136	0,0116	0,0032	0,0157	0,0141	0,0033
2012	1.463	0,0140	0,0115	0,0027	0,0146	0,0139	0,0029
2013	1.841	0,0051	0,0074	0,0020	0,0119	0,0123	0,0026
2014	1.598	0,0074	0,0063	0,0022	0,0121	0,0116	0,0027
2015	1.601	0,0036	0,0074	0,0020	0,0111	0,0123	0,0027
2016	1.367	0,0088	0,0068	0,0025	0,0118	0,0118	0,0029
2017	865	0,0098	0,0073	0,0030	0,0133	0,0120	0,0033
2018	824	0,0088	0,0192	0,0033	0,0137	0,0193	0,0035
2019	892	0,0153	0,0094	0,0029	0,0156	0,0134	0,0032
2020	812	0,0143	0,0139	0,0034	0,0151	0,0141	0,0032
2021	565	0,0147	0,0099	0,0037	0,0162	0,0140	0,0036
Total	17.287	0,0080	0,0082	0,0013	0,0108	0,0123	0,0024

Análisis del primer par de dígitos (d_1-d_2)

En el análisis del primer par de dígitos significativos, se observó un patrón similar al análisis de d_1 y d_2 . Se encontró conformidad de los datos a las primeras dos posiciones decimales entre 2009 y 2019, en contraste con lo observado al inicio y al final del período de estudio (Tabla 1; Figura 4). Mientras que para el análisis de d_1 y d_2 no se observó conformidad en solo cuatro de los 18 años

del período, para el análisis de d_1-d_2 ocurrió en ocho años, lo que indicó una mayor sensibilidad en este enfoque.

DISCUSIÓN Y CONCLUSIONES

En el ámbito de la administración y conservación de los recursos pesqueros, la confiabilidad de los

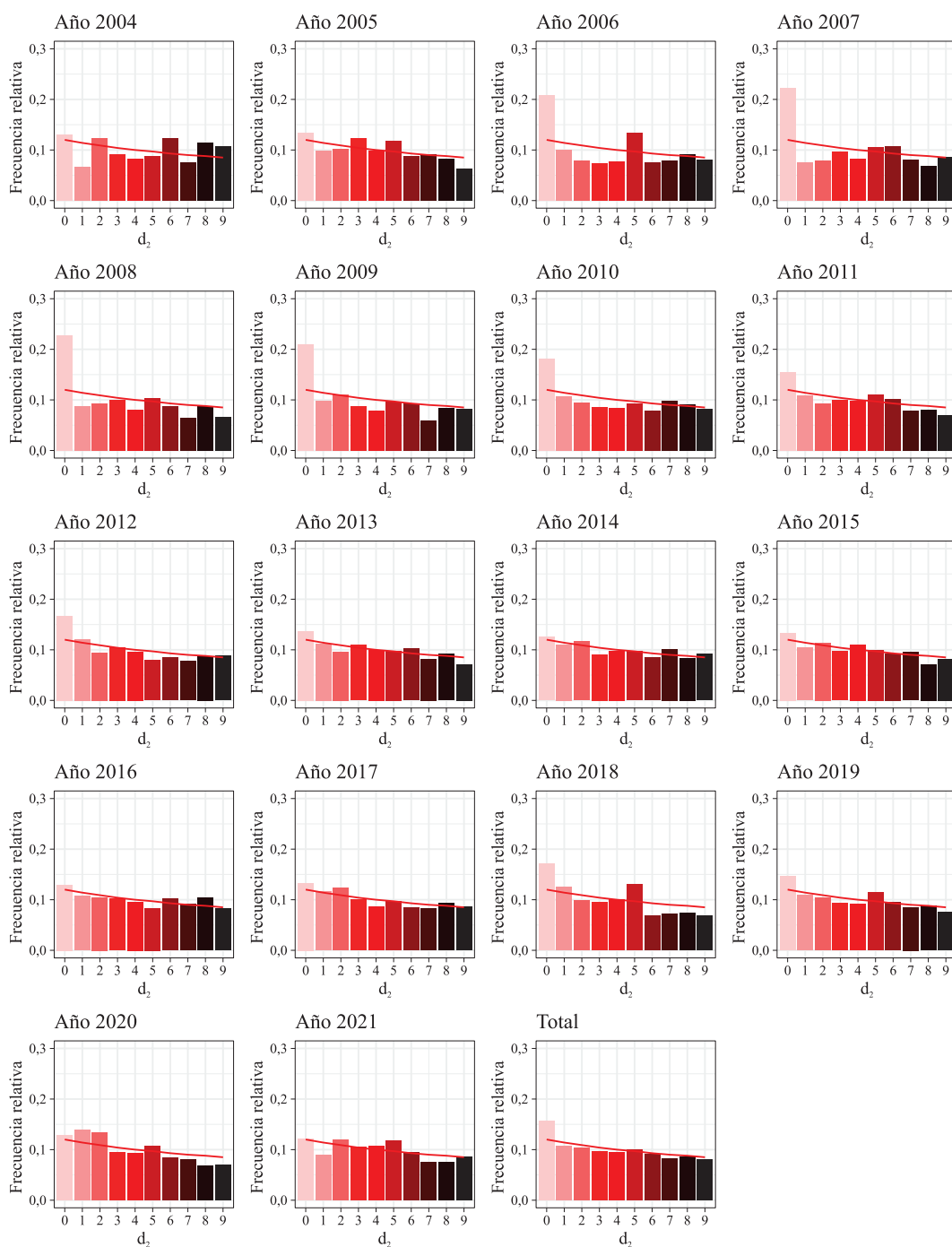


Figura 3. Distribución de frecuencia observada en el segundo dígito (d_2) de la declaración oficial de captura de corvina rubia (*Micropogonias furnieri*) en el periodo 2004-2021 (barras) y la distribución teórica esperada de Benford (línea).

Figure 3. Observed frequency distribution of the second digit (d_2) in the official catch declaration of Argentine whitemouth croaker (*Micropogonias furnieri*) during the period 2004-2021 (bars) and the expected theoretical Benford distribution (line).

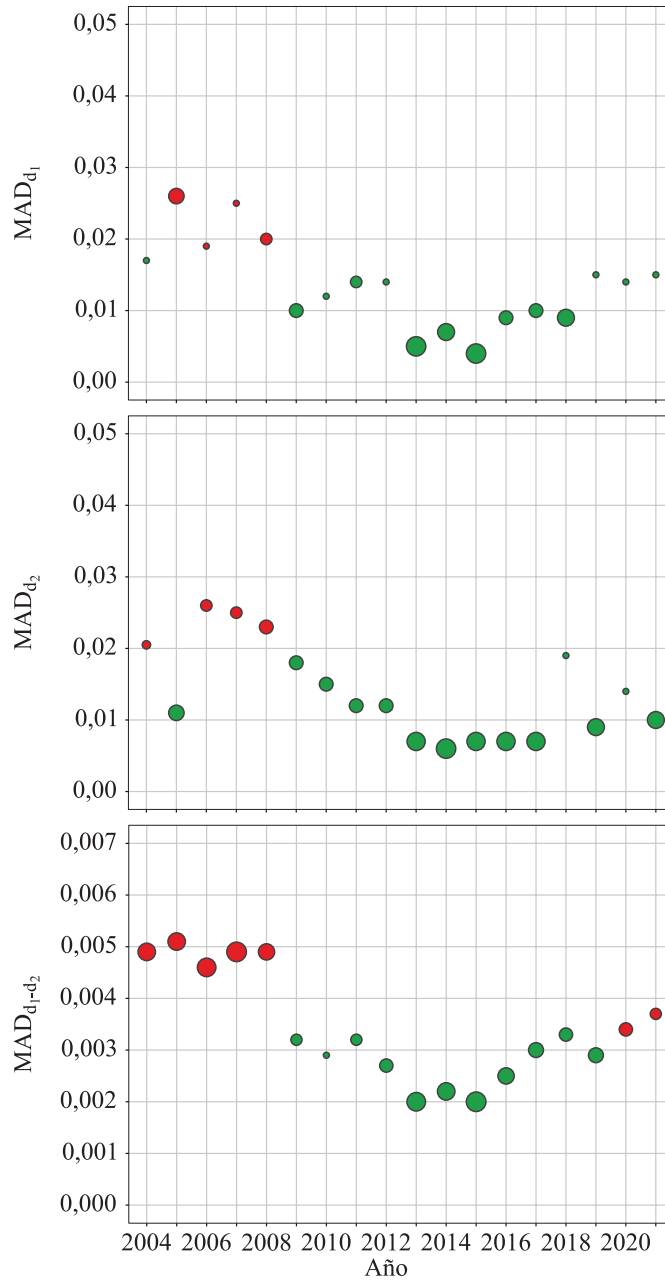


Figura 4. Tendencia anual de la prueba de conformidad de MAD para el primer dígito (d_1), segundo dígito (d_2) y el primer par de dígitos (d_1-d_2) de la captura (kg) de corvina rubia (*Micropogonias furnieri*) reportada por la flota comercial argentina en el período 2004-2021. En negritas se resaltan los valores que exceden el umbral crítico de conformidad. El tamaño de los círculos indica la magnitud de la diferencia entre los valores observados y críticos.

Figure 4. Annual trend of the MAD conformity test for the first digit (d_1), second digit (d_2), and first pair of digits (d_1-d_2) of the reported catch (kg) of Argentina whitemouth croaker (*Micropogonias furnieri*) by the Argentine commercial fleet during the 2004-2021 period. Values exceeding the critical conformity threshold are highlighted in bold. The size of circles indicates the magnitude of the difference between observed and critical values.

datos resulta esencial para respaldar el desarrollo de políticas y estrategias efectivas de manejo. En este estudio, evaluamos la calidad de la estadística oficial de captura de *M. furnieri* reportada en los partes de pesca provinciales y nacionales argentinos mediante el análisis de conformidad con la ley de Benford. Esta información es de gran relevancia dado que se utiliza como insumo en la construcción de índices de abundancia basados en la captura y el esfuerzo pesquero (García 2021; Rico 2021; Riestra 2021). Los resultados obtenidos proporcionaron una evaluación detallada de los datos de captura reportados por la flota que operó bajo la modalidad “a la pareja” desde el puerto de Mar del Plata en un período de 18 años. En este sentido, se observó conformidad en términos de la primera y segunda posición decimal, aunque en virtud de la cantidad de años en los cuales se rechazó la ley de Benford, el análisis del primer par de dígitos demostró ser el más sensible al detectar posibles irregularidades. La discriminación anual de la información permitió evaluar la calidad de los datos a lo largo del tiempo, observándose una mayor confiabilidad en los reportes de captura más recientes en comparación con los de años anteriores.

El procedimiento empleado para estimar los valores críticos de la prueba de conformidad MAD en cada año del período 2004-2021 abordó dos problemáticas fundamentales: la naturaleza de los datos y la variabilidad en el tamaño de muestra siguiendo las recomendaciones de Kössler y colaboradores (2021). Al sustituir las estimaciones previas de valores críticos para MAD, basadas en un amplio conjunto de datos genéricos (Nigrini 2012), y reemplazar un único valor fijo del mismo por estimaciones anuales o específicas de cada set de datos se intentó reducir el sesgo señalado por Cerqueti y Lupi (2021). Los valores críticos estimados en este trabajo para d_1 , d_2 y d_1-d_2 , presentaron rangos que incluyeron a los valores de referencia establecidos por Nigrini (2012), lo que sugiere que la metodología aplicada es consistente con los estándares previamente establecidos. Resultaría importante complementar el análisis de la conformidad con

la ley de Benford con otros métodos estadísticos para una evaluación más robusta de la calidad de los datos (Cerri 2018). La aplicación de un enfoque multidimensional que incluya pruebas adicionales podría proporcionar una perspectiva más completa y mitigar las limitaciones inherentes a cualquier método individual. La implementación de técnicas de aprendizaje automático como redes neuronales artificiales, árboles de decisión, *random forest*, etc., permitirían profundizar el análisis al identificar patrones complejos y sutiles que no resulten tan evidentes (Badal Valero et al. 2018). Futuras investigaciones deberían explorar la integración de estas herramientas para fortalecer la auditoría de datos en la conservación y gestión de recursos naturales.

AGRADECIMIENTOS

Agradecemos enormemente a Hans-J Lenz, del Instituto de Estadística y Econometría de la Universidad Libre de Berlín, Alemania, por compartir información sobre las simulaciones utilizadas en Kössler et al. (2021) y por sus comentarios, los cuales permitieron enriquecer nuestro trabajo. Asimismo, extendemos nuestro agradecimiento a los revisores anónimos cuyos comentarios constructivos han permitido mejorar sustancialmente este artículo, enriqueciendo significativamente el contenido y la claridad del mismo. Contribución INIDEP N° 2362.

Contribución de autores

Sebastián García: conceptualización, validación, investigación, curación de datos, escritura-borrador original, escritura-revisión y edición, visualización. Julieta S. Rodríguez: conceptualización, metodología, *software*, validación, análisis formal, investigación, curación de datos, escritura-borrador original, escritura-revisión y edición. Bruno V. Menna: metodología, *software*, análisis formal,

curación de datos, escritura-borrador original, escritura-revisión y edición.

REFERENCIAS

- BADAL-VALERO E, ALVAREZ-JAREÑO JA, PAVÍA JM. 2018. Combining Benford's law and machine learning to detect money laundering. An actual Spanish court case. *Forensic Sci Int.* 282: 24-34. DOI: <https://doi.org/10.1016/j.forsciint.2017.11.008>
- BARNEY B, SCHULZKE K. 2016. Moderating "cry wolf" events with excess MAD on Benford's law research and practice. *J Forensic Accounting Res.* 1 (1): 66-90.
- BENFORD F. 1938. The law of anomalous numbers. *Proc Am Philos Soc.* 78: 551-572.
- BLITZSTEIN JK, HWANG J. 2020. Benford's Law: theory, applications, and limitations. En: LEE C-F, LEE JC, editores. *Handbook of financial econometrics and statistics.* Nueva York: Springer. 1-27.
- CABEZA GARCÍA PM. 2019. Aplicación de la ley de Benford en la detección de fraudes. *Univ Soc.* 11 (5): 421-427.
- CAROZZA C, RUARTE CO, RICO MR, LAGOS AN, GARCÍA S, RIESTRA C, LORENZO MI. 2019. La pesquería del variado costero. Evolución de los desembarques y recomendación de la Captura Biológicamente Aceptable efectuadas a la CTM-FM para las principales especies costeras óseas. Año 2018. *Inf Téc INIDEP N° 5/2019.* 62 p.
- CAROZZA C, NAVARRO L, JAUREGUIZAR A, LASTA C, BERTOLOTTI MB. 2001. Asociación íctica costera bonaerense "variado costero". *Inf Téc Int DNI-INIDEP N° 38/2001.*
- CERQUETI R, LUPI C. 2021. Some new tests of conformity with Benford's Law. *Stats.* 4 (3): 745-761. DOI: <https://doi.org/10.3390/stats4030044>
- CERRI J. 2018. A fish rots from the head down: how to use the leading 2 digits of ecological data to detect their falsification. DOI: <https://doi.org/10.1101/368951>
- CINELLI C. 2014. Benford. Analysis: Benford analysis for data validation and forensic analytics. R package version 0.1. 1.
- DIEKMANN A. 2007. Not the first digit! Using Benford's Law to detect fraudulent scientific data. *J Appl Stat.* 34: 321-329.
- GARCÍA S. 2023. La pesquería comercial argentina de corvina rubia (*Micropogonias furnieri*) entre los 34° S y los 39° S. Año 2021. *Inf Invest INIDEP N° 4/2023.* 11 p.
- GARCÍA S, MARTINEZ PULJAK G, HERNÁNDEZ, D. 2018. Uso del monitoreo satelital como indicador del esfuerzo pesquero de la flota comercial argentina. *Inf Invest INIDEP N° 26/2018.* 19 p.
- KÖSSLER W, LENZ HJ, WANG XD. 2021. Is the Benford law useful for data quality assessment? En: KNOTH S, SCHMID W, editores. *Frontiers in statistical quality control 13.* ISQC 2019. Cham: Springer. DOI: https://doi.org/10.1007/978-3-030-67856-2_22
- MAUNDER M, PUNT A. 2004. Standardizing catch and effort data: a review of recent approaches. *Fish Res.* 70. 141-159
- MEBANE W. 2008. Election forensics: the second digit Benford's Law test and recent American presidential elections. Detecting and deterring electoral manipulation. Washington: Brooking Press.
- MOCHTY L. 2002. Die Aufdeckung von Manipulationen im Rechnungswesen - was leistet das Benford's Law. *Die Wirtschaftsprüfung.* 55 (14): 725-736.
- MORROW J. 2014. Benford's Law, families of distributions and a test basis. *CEP Discussion Papers N° 1291.* Londres: Centre for Economic Performance.
- NEWCOMB S. 1881. Note on the frequency of use of different digits in natural numbers. *Am J Math.* 4 (1): 39-40.
- NIGRINI M. 1999. I've got your number. *J Account.* 187: 79-83.
- NIGRINI M. 2012. Benford's Law: applications for forensic accounting, auditing, and fraud detec-

- tion. Vol. 586. John Wiley Sons.
- NOLETO-FILHO EM, CARVALHO AR, THOMÉ-SOUZA MJF, ANGELINI R. 2022. Reporting the accuracy of small-scale fishing data by simply applying Benford's law. *Front Mar Sci.* 9: 947503. DOI: <https://doi.org/10.3389/fmars.2022.947503>
- PAULY D, BELHABIB D, BLOMEYER R, CHEUNG WL, CISNEROS-MONTEMAYOR AM, COPELAND D, HARPER S. 2014. China's distant-water fisheries in the 21st Century. *Fish Fish.* 15: 474-488.
- PEARSON K. 1900. 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. *Philos Mag Ser. (5)* 50: 157-175.
- PINKHAM, R. 1961. On the distribution of the first significant digits. *Ann Math Stat.* 32: 1223-1230.
- R DEVELOPMENT CORE TEAM. 2019. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- SCHISTERMAN EF, PERKINS NJ, LIU A, BONDELL H. 2005. Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology.* (16) 1: 73-81. DOI: <https://doi.org/10.1097/01.ede.0000147512.81966.ba>
- SCHRÄPLER Jörg P. 2010. Benford's Law as an instrument for fraud detection in surveys using the data of the socio-economic panel (SOEP). *SOEPpapers.* 273. 56 p. DOI: <https://doi.org/10.2139/ssrn.1562574>
- SILVA L, FILHO DF. 2021. Using Benford's Law to assess the quality of COVID-19 register data in Brazil. *J Pubic Health.* 43: 107-110.
- SCOTT P, FASLI M. 2001. Benford's Law: an empirical investigation and a novel explanation. CSM Technical Report. 349. Department of Computer Science, University Essex.
- TSAGBEY S, DE CARVALHO M, PAGE GL. 2017. All data are wrong, but some are useful? Advocating the need for data auditing. *Am Stat.* 71 (3): 231-235.
- TESFAMICHAEL D, PAULY D. 2011. Learning from the past for future policy: approaches to time series catch data reconstruction. *West Indian Ocean J Mar Sci.* 10 (2): 99-106.

