Revista de Psicología, 43(2), 2025, pp. 621-654 (e-ISSN 2223-3733)

Estimación por intervalo de la Razón de Variación Universal

José Moral de la Rubia
Universidad Autónoma de Nuevo León, México

Las variables cualitativas son de gran importancia en las ciencias sociales y se han desarrollado medidas descriptivas de variabilidad y forma para estas variables, pero se utilizan poco. Una de las principales razones es que no están disponibles en los programas estadísticos. Este artículo se centra en una medida de variabilidad recientemente introducida, la Razón de Variación Universal, que mejora las razones de variación previamente formuladas. El objetivo del artículo es plantear el cálculo de su error estándar e intervalo de confianza y proporcionar un guion para su cómputo usando el programa R. Se aplica a dos escenarios: con una moda y con dos o más modas. Se recomienda integrar esta medida de variabilidad en procedimientos de análisis de datos.
Palabras clave: razón de variación, variables cualitativas, intervalo de confianza, remuestreo con reemplazamiento, método de percentil corregido de sesgo y acelerado.

Interval estimation of Universal Variation Ratio
Qualitative variables are of great importance in the social sciences and descriptive measures of variability and shape have been developed for these variables, but they are little used. One of the main reasons is that they are not available in statistical software. This paper focuses on a recently introduced measure of variability, the Universal Variation Ratio, which enhances upon variation ratios previously formulated. This article aims to propose the calculation of its standard error and confidence interval and to provide a script for its computation using the R program. It applies to two scenarios: with one mode and with two or more modes. It is recommended to integrate this variability measure into data analysis procedures.
Keywords: variation ratio, qualitative variables, confidence interval, bootstrap, bias-corrected and accelerated percentile method.

Estimativa intervalar do Rácio de Variação Universal
As variáveis qualitativas são de grande importância nas ciências sociais e foram desenvolvidas medidas descritivas de variabilidade e forma para estas variáveis, mas são pouco utilizadas. Uma das principais razões é o facto de não estarem disponíveis em software estatístico. Este artigo centra-se numa medida de variabilidade recentemente introduzida, o Rácio de Variância Universal, que melhora os rácios de variância anteriormente formulados. O objetivo deste artigo é propor o cálculo do seu erro padrão e intervalo de confiança e fornecer um script para o seu cálculo utilizando o programa R. Aplica-se a dois cenários: com um modo e com dois ou mais modos. Recomenda-se a integração desta medida de variabilidade nos procedimentos de análise de dados.
Palavras chave: rácio de variação, variáveis qualitativas, intervalo de confiança, reamostragem com reposição, correção de viés e método do percentil acelerado.

José Moral de la Rubia http://orcid.org/0000-0003-1856-1458
Toda correspondencia concerniente a este artículo debe ser dirigido a José Moral de la Rubia. Email: jose.morald@uanl.edu.mx

https://doi.org/10.18800/psico.202502.001

Se entiende por variable cualitativa un criterio de clasificación (definición intensional) que permite identificar y distinguir un conjunto numerable de categorías nominales o atributos. La especificación completa de este conjunto constituye la definición extensional de la variable. Las categorías nominales se pueden representar por números, pero estos números no permiten operaciones aritméticas. Son intercambiables por símbolos o palabras y solo permiten contar frecuencia o contingencia. Las variables cualitativas se suelen representar por letras latinas mayúsculas (A, B o C) y sus k categorías nominales se pueden denotar con la letra de la variable en minúscula con un subíndice. Ejemplos de variables cualitativas son: A = identidad sexual autodefinida y no disfórica = {a_1 = cisgénero femenina o mujer, a_2 = cisgénero masculina u hombre, a_3 = hombre transexual, a_3 = hombre transgénero, a_5 = mujer transexual, a_5 = mujer transgénero, a_7 = no binaria) o B = preferencia sexual no parafílica = {b_1 = asexual, b_2 = bisexual, b_3 = fluida, b_4 = heterosexual, b_5 = homosexual, b_6 = pansexual} (Hammack, Hughes, Atwood, Cohen & Clark, 2022).

Se han definido muchas medidas de variación para este tipo de variables (Moral-de la Rubia, 2022a; Wilcox, 1973), sin embargo, son poco conocidas y utilizadas, aun cuando las variables cualitativas son frecuentes y muy importantes en la investigación social y de la salud (Levitt, 2021; Maxwell, 2021). De hecho, en muchos manuales de estadística básica o aplicada ni siquiera se mencionan y los programas estadísticos comerciales o de acceso libre no las incluyen. Es importante destacar que su desarrollo es relativamente reciente, comenzando en la mitad del siglo XX con la publicación del artículo del matemático estadounidense Claude Elwood Shannon (1916–2001) sobre la teoría matemática de la comunicación (Golan & Harte, 2022), de donde surge la entropía estandarizada o relativa (Shannon, 1948), y con la propuesta del índice de diversidad por Simpson (1949). Estas medidas

proliferaron entre las décadas de 1960 a 1980, destacando entre ellas la razón de variación de Freeman (1965), la razón de variación de la moda de Wilcox (1973), el índice de variación cualitativa de Gibbs y Poston (1975) y la desviación estándar desde la moda de Kvalseth (1988). No obstante, aún hoy en día se siguen desarrollando (Evren & Erhan, 2017; Li, Garg & Deng, 2020; Moral-de la Rubia, 2022a; Weiss, 2019).

Una característica que comparten todas estas medidas de variabilidad para variables cualitativas es poseer un rango de 0 a 1. En este rango, el 0 corresponde a la distribución de una variable aleatoria constante, en la que un valor concentra toda la probabilidad a nivel poblacional o toda la frecuencia a nivel muestral. El valor de 1 corresponde a una distribución uniforme, en la que todos sus valores tienen la misma probabilidad a nivel poblacional o la misma frecuencia a nivel muestral. Estas medidas son de cálculo sencillo, prácticas para su uso convencional, de interpretación clara y se han aplicado en investigación social, pero con poca frecuencia (Deacon & Stanyer, 2021; Evren, Tuna, Ustaoglu & Sahin, 2021; Moral-de la Rubia, 2022a). Una de las principales razones de su infrautilización es que no se incluyen en los programas estadísticos u hojas de cálculo de uso común, ya sean comerciales, Excel (Microsoft, 2021), SPSS (IBM Documentation, 2023) o STATA (StataCorp, 2023) o de libre acceso, como JACS (Goss-Sampson, 2020); R (R Foundation, 2024) o Real Statistics using Excel (Zaiontz, 2024). Incluso no aparecen en programas específicos de análisis cualitativo que permiten generar variables de categorías nominales y ordenadas, como ATLAS.ti (ATLAS.ti Scientific Software Development GmbH., 2023) o MAXQDA (VERBI Software, 2024).

Este trabajo metodológico se centra en una medida de variabilidad recientemente introducida por Moral-de la Rubia (2022a), la Razón de Variación Universal, que mejora las razones de variación previamente formuladas por Freeman (1965) y Wilcox (1973). El objetivo del artículo es plantear el cálculo del error estándar e intervalo de confianza del estadístico, así como proporcionar un guion de códigos para su cómputo usando el programa R que es de acceso gratuito (R Foundation, 2024).

Para este segundo objetivo, se consideran dos situaciones de cálculo: muestras aleatorias de datos cualitativos con una sola moda y con dos o más modas. Ambas muestras fueron generadas, aunque se le da un contenido específico para que resultan más familiares y comprensibles para los lectores.

Razón de variación universal

La razón de variación de Freeman (1965) parte de una fórmula de variación en torno a la moda y se simplifica al complemento de la frecuencia de la moda: $RVF = 1 - f_{mo}$. Se aplica únicamente en casos de distribuciones unimodales o en el caso de que la distribución no tenga moda (uniforme), argumentando que la frecuencia modal es nula ($f_{mo} = 0$). Si la muestra o la población presenta dos o más modas, esta medida de variación no puede ser calculada.

Wilcox (1973) introdujo una razón de variación estandarizada basada en la moda, algo más sofisticada que incluía información sobre el número de categorías nominales: $RVW = 1 - (k \times f_{mo} - 1) / (k - 1)$. Ambas medidas se relacionan: $RVW = RVF \times k / (k - 1)$. Con base en esta relación, se puede deducir que la razón de variación de Wilcox es mayor o igual que la del Freeman: $RVW \geq RVF$. Esta medida de variación requiere necesariamente que la distribución sea unimodal. Cuando hay dos o más modas, no se puede calcular. En caso de una distribución uniforme, no se puede dar a la frecuencia modal un valor de 0, como se argumentó con la razón de variación de Freeman (1965), ya que *RVW* queda fuera del rango de 0 a 1 estipulado para los índices de variación estandarizados de variables cualitativas.

Moral-de la Rubia (2022a) planteó una modificación de la fórmula de Freeman (1965). Esta nueva propuesta permite su aplicación en casos de múltiples modas y considera el número de categorías (k), al igual que hace la razón de variación de la moda (*RVW*) de Wilcox (1973). El autor denominó a este estadístico modificado como razón de variación universal (*RVU*), debido a que puede ser aplicado con cualquier tipo de distribución de variable cualitativa. Véase la Fórmula 1. En esta fórmula, k representa el número de categorías cualitativas de la

variable; c es el número de valores con frecuencia (absoluta o relativa) máxima, cuyo valor puede variar de 1 a k; y f_{max} denota la frecuencia relativa máxima que corresponde a la moda (f_{mo}), salvo en una distribución uniforme ($c = k$) en la cual se considera que no hay moda y el valor de todas las frecuencias es $1/k$.

$$RVU = \frac{1-\frac{1}{c}\times f_{max}}{1-max\left(\frac{1}{c}\times f_{max}\right)} = \frac{1-\frac{1}{c}\times f_{max}}{1-\frac{1}{k^2}} = \frac{1-\frac{1}{c}\times f_{max}}{\frac{k^2-1}{k^2}}$$
$$= \frac{k^2}{k^2-1}\times\left(1-\frac{f_{max}}{c}\right) \qquad (1)$$

Partiendo de la fórmula de Freeman (1965), $1 - f_{mo}$, la fórmula propuesta pondera la frecuencia relativa modal por el inverso del número de modas ($1 / c$) y divide la expresión por su valor máximo. Este máximo se alcanza con la distribución uniforme, cuando $c = k$, $f_{max} = 1 / k$ y $1 - 1 / c \times f_{max} = 1 - 1 / k \times 1 / k = 1 - 1 / k^2$. Cuando un valor acapara toda la frecuencia (variable aleatoria constante), donde la frecuencia modal es única y tiene un valor unitario ($c = 1$ y $f_{max} = 1$), la razón de variación universal (*RVU*) alcanza su valor mínimo de 0. Como se puede ver en la Fórmula 2.

$$RVU = \frac{k^2}{k^2-1}\times\left(1-\frac{f_{max}}{c}\right) = \frac{k^2}{k^2-1}\times\left(1-\frac{1}{1}\right) = \frac{k^2}{k^2-1}\times 0 = 0 \qquad (2)$$

Si no hay moda (distribución uniforme), todas las categorías tienen la misma frecuencia y dicha frecuencia es la máxima ($1 / k$), el valor de c es k y la razón de variación universal (*RVU*) alcanza su valor máximo de 1. Véase Fórmula 3.

$$RVU = \frac{1-\frac{f_{max}}{c}}{\frac{k^2-1}{k^2}} = \frac{1-\frac{1/k}{k}}{\frac{k^2-1}{k^2}} = \frac{1-\frac{1}{k^2}}{1-\frac{1}{k^2}} = 1 \qquad (3)$$

En la medida en que c se aproxima a k, siendo k el número de categorías (distribución limitante uniforme para la cota superior), el resultado de la modificación propuesta se aproxima a 1 (Fórmula 4).

$$RVU = \lim_{c \to k} \frac{1 - \frac{1}{c} \times f_{max}}{1 - \frac{1}{k^2}} = \frac{1 - \frac{1}{k} \times \frac{1}{k}}{1 - \frac{1}{k^2}} = \frac{1 - \frac{1}{k^2}}{1 - \frac{1}{k^2}} = 1 \qquad (4)$$

Esto se debe a que, en la medida en que la muestra de la variable cualitativa A presenta más categorías con frecuencia máxima (c), el efecto sustractor de la frecuencia máxima disminuye en la razón de variación modificada ($1 - f_{max} / c$) y, consecuentemente, el valor de esta medida de variación aumenta (*RVU*). Cuanto menor es el número de categorías (k), mayor es el incremento en la razón de variación universal (*RVU*), ya que se reparte más la variabilidad, alejándose la distribución de la variable A de la de una variable aleatoria constante (valor mínimo) y aproximándose más a una distribución uniforme (valor máximo).

En el caso de una moda (c = 1), que es la situación en que la fórmula de Freeman (1965) y *RVU* son comparables, la *RVU* arroja a un valor mayor o igual que la *RV* de Freeman (1965). Véase Fórmula 5.

$$RVU = \frac{1 - \frac{1}{c} f_{max}}{\frac{k^2 - 1}{k^2}} = \frac{1 - \frac{1}{1} f_{mo}}{\frac{k^2 - 1}{k^2}} = \frac{k^2}{k^2 - 1}\left(1 - f_{mo}\right)$$

$$= \frac{k^2}{k^2 - 1} RVF \geq RVF = 1 - f_{mo} \qquad (5)$$

Cuando el número de categorías cualitativas (k) es muy pequeño, hay mayor diferencia entre *RVU* y *RVF*. Con dos categorías, la diferencia o incremento es de un tercio: $RVU - RVF = k^2 / (k^2 - 1) = 0.333$. Con tres categorías, la diferencia o incremento es de un octavo: $RVU - RVF = k^2 / (k^2 - 1) = 0.125$. No obstante, en la medida que se incrementa el número de categorías, la diferencia es menor. Con cuatro categorías,

la diferencia o incremento es de 0.067 y, con cinco, de 0.042. Cuando el número de categorías tiende a infinito, *RVU* converge a *RVF*. Véase la Fórmula 6.

$$Si\, c = 1, RVU = \lim_{k\to\infty} \frac{k^2}{k^2-1}\left(1-\frac{f_{max}}{c}\right) = \lim_{k\to\infty} \frac{k^2}{k^2-1}\left(1-f_{mo}\right)$$
$$= 1 - f_{mo} = RVF \tag{6}$$

Cuando hay una única categoría cualitativa con frecuencia máxima (c = 1), la nueva medida de variación propuesta por Moral-de la Rubia (2022a), la razón de variación universal (*RVU*), es menor o igual que la razón de variación de la moda (*RVW*) de Wilcox (1973). Consecuentemente, *RVF* siempre toma un valor menor o igual que *RVU* y *RVU* siempre toma un valor menor o igual que *RVW*. Véase Fórmula 7.

$$RVU = \frac{k^2}{k^2-1} \times RVF \rightarrow RVF = \frac{k^2-1}{k^2} \times RVU = \left(1-\frac{1}{k^2}\right) \times RVU$$
$$RVW = \frac{k}{k-1} \times RVF \rightarrow RVF = \frac{k-1}{k} \times RVW = \left(1-\frac{1}{k}\right) \times RVW$$
$$RVF = \left(1-\frac{1}{k^2}\right) \times RVU = \left(1-\frac{1}{k}\right) \times RVW$$
$$RVF \leq RVU \leq RVW \tag{7}$$

Propuesta para el cálculo del intervalo de confianza

Dado que la distribución en el muestreo de *RVU* es desconocida, se puede recurrir al muestreo repetitivo con reemplazamiento para obtenerla (Rousselet, Pernet & Wilcox, 2023). A partir de dicha distribución, a través del método no paramétrico de percentil o el método de percentil corregido de sesgo y asimetría (acelerado), se puede definir el intervalo de confianza (Zelikman, Wu, Mu & Goodman, 2022). En caso de presentarse sesgo y asimetría, se prefiere el segundo método

sobre el primero (Rousselet, Pernet & Wilcox, 2021). Si este segundo método arroja un percentil corregido de sesgo con un valor extremo, impidiendo el cálculo del intervalo de confianza, se puede asignar un valor de 0 a dicho percentil, lo que proporciona un resultado equivalente al método de percentil, como sugiere Efron (1987). A continuación, se presentan los algoritmos de cálculo para obtener el intervalo de confianza mediante ambos métodos (Canty, 2022).

1. Se parte de una muestra aleatoria de tamaño *n* o vector *n* dimensional de la variable D: d = $\{d_1, d_2, \ldots, d_n\} \subseteq$ D. La variable D puede ser cualitativa, ordinal o cuantitativa. En el presente caso, el foco de interés recae sobre las variables cualitativas.

2. Se calcula el estadístico o estimador: $\hat{\theta} = t(\mathrm{x})$, lo que constituye la estimación del parámetro θ en la muestra original de tamaño *n*. En el presente caso, el estadístico es la razón de variación universal (*RVU*).

3. Se extraen aleatoriamente y con reemplazamiento *B* muestras de tamaño *n* de la muestra aleatoria original de tamaño *n*. Se recomienda que sean al menos 1000 extracciones ($B \geq 1000$), que el tamaño de la muestra sea de al menos 30 ($n \geq 30$) y que la muestra aleatoria sea representativa de la población de la que se recolectó. Este método es inadecuado con muestras pequeñas y no aleatorias (Rousselet et al., 2021).

4. En cada una de las *B* muestras, se calcula la medida de variación: $\hat{\theta}_i^*$ ($i = 1, 2, \ldots, B$), con lo que se genera la denominada *distribución en el muestreo Bootstrap del estadístico o estimador*. Esta distribución se puede representar por medio de un histograma con la curva de densidad superpuesta. La amplitud o ancho uniforme (*w*) y el número de intervalos de clase o rectángulos (*k*) del histograma se pueden definir por la regla de Freedman y Diaconis (1981), que no requiere ningún supuesto distribucional y proporciona un histograma óptimo (Contreras-Reyes & Brito, 2022).

5. La Fórmula 8 muestra el cálculo del ancho uniforme (*w*) por esta regla, donde R_{IC} es el rango intercuartílico, C_3 es el cuartil superior y C_1 es el cuartil inferior de la muestra Bootstrap con *B* datos. A su vez, la Fórmula 9 muestra el cálculo del número de intervalos de clase o rectángulos (*k*), donde *R* es el rango, *max* es el valor más alto y *min* es el valor más bajo en la distribución en el muestreo Bootstrap del estadístico o estimador que contiene *B* datos.

$$w = \frac{2 \times R_{IC}\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right)}{\sqrt[3]{B}} = \frac{2 \times \left(C_3\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right) - C_1\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right)\right)}{\sqrt[3]{B}} \tag{8}$$

$$k = \frac{R\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right)}{w} = \frac{max\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right) - min\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right)}{w} \tag{9}$$

6. La estimación Bootstrap del parámetro θ es la media aritmética de la distribución en el muestreo Bootstrap del estadístico o estimador y se denota por $\hat{\theta}_{bootstrap}$ (Fórmula 10). El error estándar Bootstrap es la desviación estándar muestral o corregida de sesgo de la distribución en el muestreo Bootstrap del estadístico o estimador y se denota por $ee_{boostrap}$ (Fórmula 11) El sesgo Bootstrap es la diferencia entre la estimación Bootstrap y la estimación en la muestra original ($\hat{\theta}$) y se denota por $sesgo_{boostrap}$ (Fórmula 12).

$$\hat{\theta}_{bootstrap} = \overline{\hat{\theta}}^* = \frac{\sum_{i=1}^{B} \hat{\theta}_i^*}{B} \tag{10}$$

$$ee_{bootstrap} = \sqrt{\frac{\sum_{i=1}^{B}\left(\hat{\theta}_i^* - \hat{\theta}_{bootstrap}\right)^2}{B-1}} \tag{11}$$

$$sesgo_{boostrap} = \hat{\theta}_{bootstrap} - \hat{\theta} \tag{12}$$

El cuantil de orden α / 2 ($c_{\alpha/2}$) de la distribución en el muestreo Bootstrap del estadístico o estimador define el límite inferior y el cuantil de orden 1 – α /2 ($c_{1-\alpha/2}$) establece el límite superior del intervalo de confianza al $(1 - \alpha) \times 100$ por el método percentil (Fórmula 13).

$$P\left(c_{\alpha/2}\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right) \leq \theta \leq c_{1-\alpha/2}\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right)\right) = 1-\alpha \qquad (13)$$

El cuantil se puede obtener por la regla 8 (de interpolación lineal) del programa R, como Hyndman y Fan (1996) recomiendan para el cálculo de cuantiles muestrales. Esta regla se basa en la mediana del estadístico de orden i de una muestra de tamaño B extraída de una distribución uniforme estándar o distribución beta con sus dos parámetros de forma unitarios: $U[0, 1] \equiv \text{Beta}(\alpha = 1, \beta = 1)$. Esta distribución continua permite modelar la situación cuando un valor de probabilidad es desconocido (prior no informado en estimación bayesiana). Aunque el orden del cuantil o probabilidad acumulada p no es propiamente desconocido, sí lo es el orden i en la muestra de tamaño B del cuantil de orden p, siendo i la incógnita que se despeja. La distribución en el muestreo del estadístico de orden i de una muestra de tamaño B de una distribución uniforme estándar es una distribución beta de parámetros de forma: $\alpha = i \geq 1$ y $\beta = B + 1 - i \geq 1$ (cuando $\alpha = \beta > 1$). Cuando α y $\beta > 1$ o $\alpha \neq \beta \geq 1$, la mediana de la distribución beta es: $Mdn(\text{X}) \approx (\alpha - 1/3) / (\alpha + \beta - 2/3) = (i - 1/3) / (i + B + 1 - i - 2/3) = (i - 1/3) / (B + 1/3)$.

Con la regla 8, se expresa el orden del cuantil como: $p = (i - 1/3) / (B + 1/3)$, donde i es el orden desconocido del dato al que le corresponde el cuantil de orden p, una vez que los B datos muestrales de la variable cuantitativa han sido ordenados en sentido ascendente: $\hat{\theta}_{(i)}^*$. Al despejar dicha incógnita: $i = 1/3 + p \times (n + 1/3)$, se busca el dato muestral en el orden i entre los

B datos ordenados ascendentemente. Si *i* resulta un entero, el dato muestral en el orden *i* es el cuantil de orden *p* (c_p) Si *i* es un número con decimales, el cuantil se obtiene por interpolación lineal mediante la Fórmula 14. Cabe señalar que el estadístico *RVU* es una variable cuantitativa continua con dominio en el intervalo [0, 1].

$$i = \frac{1}{3} + p \times \left(B + \frac{1}{3} \right); c_p \left(\left\{ \hat{\theta}_i^* \right\}_{i=1}^{B} \right) = \hat{\theta}_{(i)}^* = \hat{\theta}_{(i)}^* + (i - i) \times \left(\hat{\theta}_{(i+1)}^* - \hat{\theta}_{(i)}^* \right) \quad (14)$$

7. Para obtener el intervalo de confianza por el segundo método, se inicia calculando el percentil corregido de sesgo, que se denota por z_0^*. Véase la Fórmula 15, donde *I* es la función indicatriz (0 cuando no se cumple la condición y 1 cuando se cumple) y Φ^{-1} es la función probit o función cuantil de la distribución normal estándar.

$$z_0^* = \Phi^{-1} \left(\frac{\sum_{i=1}^{B} I\left(\hat{\theta}_i^* \leq \hat{\theta} \right)}{B} \right) \quad (15)$$

En caso de valores extremos, cuando el argumento de la función probit se aproxima a 0 o 1, z_0^* queda indefinido. En este caso, se puede dar a z_0^* un valor de 0, con lo que el intervalo de confianza Bootstrap corregido por sesgo y acelerado (PCSA) es muy semejante al intervalo de confianza Bootstrap por el método percentil (Efron, 1987).

8. Se calcula el factor de corrección de asimetría (aceleración), utilizando el método de la navaja ("jackknife" en inglés), esto es, generando *n* muestras a partir de la muestra aleatoria d de tamaño *n*, eliminando un dato muestral en cada muestra, y calculando el estadístico o estimador en cada una de las *n* muestras $\hat{\theta}_{(-i)}$, como se muestra en la Fórmula 16. Antes de calcular el factor de aceleración, denotado por *a*, se requiere

obtener la estimación Jackknife del parámetro a través de la media de la *distribución en el muestreo Jackknife del estimador* (Fórmula 17).

$$a = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{Jackknife} - \hat{\theta}_{(-i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\theta}_{Jackknife} - \hat{\theta}_{(-i)}\right)^2\right]^{3/2}} \tag{16}$$

$$\hat{\theta}_{Jackknife} = \overline{\hat{\theta}}_{(-i)} = \frac{\sum_{i=1}^{n}\hat{\theta}_{(-i)}}{n} \tag{17}$$

9. Se obtienen los órdenes de los cuantiles corregidos de sesgo y acelerados (PCSA) correspondientes al límite inferior (*LI*) y superior (*LS*) del intervalo de confianza al $(1 - \alpha) \times 100$. Véase la Fórmula 18, donde Φ es la función de distribución acumulativa normal estándar, $z_{\alpha/2}$ es el cuantil de orden $\alpha/2$ de una distribución normal estándar y $z_{1-\alpha}$ es el cuantil de orden $1 - \alpha/2$ de dicha distribución.

$$p_{LI} = \Phi\left(z_0^* + \frac{z_0^* + z_{\alpha/2}}{1 - a\left(z_0^* + z_{\alpha/2}\right)}\right) \text{ y } p_{LS} = \Phi\left(z_0^* + \frac{z_0^* + z_{1-\alpha/2}}{1 - a\left(z_0^* + z_{1-\alpha/2}\right)}\right) \tag{18}$$

10. El cuantil de orden p_{LI} de la distribución en el muestreo Bootstrap del estadístico o estimador define el límite inferior y el cuantil de orden p_{LS} constituye el límite superior del intervalo de confianza al $(1 - \alpha) \times 100$ por el método de percentil corregido de sesgo y acelerado (PCSA), como se muestra en la Fórmula 19. Estos cuantiles se pueden computar por la regla 8 (de interpolación lineal) del programa R (Fórmula 14).

$$P\left(c_{p_{LI}}\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right) \le \theta \le c_{p_{LS}}\left(\left\{\hat{\theta}_i^*\right\}_{i=1}^{B}\right)\right) = 1 - \alpha \tag{19}$$

Cálculo de la razón de variación universal para variables cualitativas con el programa R

Una de las principales razones de la infrautilización de las medidas de variación para variables cualitativas es que no están disponibles en los programas estadísticos convencionales, ni en los programas específicos de análisis de datos cualitativos. A continuación, se muestra un guion de instrucciones para el programa R, el cual se puede adaptar a otros datos muestrales distintos a los presentados. A tal fin debe cambiarse la definición del vector de datos, el nombre de la tabla (identidad_sexual <- names(tabla)) y la instrucción "cat" del apartado de "Tabla de frecuencias", la referencia a Identidad_Sexual en el marco de datos y la etiqueta del eje x en el diagrama de barras, además de ajustarse el número de categorías en la instrucción "levels = 1:7", que aparece varias veces en el guion, al igual que el nombre de la tabla. Si hay seis categorías, levels pasaría a ser 1:6. Estas partes se destacan en azul dentro del guion.

Por una parte, el guion se aplica a una muestra aleatoria con una moda y, por otra parte, a una muestra aleatoria con dos modas. El guion incluye el cálculo del estadístico *RVU* y su intervalo de confianza al 95% por muestreo repetitivo con reemplazamiento (con 1000 extracciones) mediante dos métodos: percentil (PERC) y percentil corregido de sesgo y acelerado (PCSA). Aparte se representa la distribución Bootstrap por medio de un histograma con la curva de densidad superpuesta y se calcula el sesgo y error Bootstrap. Cabe señalar que se pueden usar tildes en la redacción de las etiquetas (cat) de los resultados, pero no al denominar variables. Además, el programa es sensible al uso de minúsculas y mayúsculas.

Ejemplo 1 de una muestra aleatoria con una moda

De la variable cualitativa A = "identidad sexual entre activistas de organizaciones no gubernamentales para la defensa de los derechos de las minorías sexuales" = {a_1 = cisgénero femenina o mujer, a_2 = cisgénero masculina u hombre, a_3 = hombre transexual, a_4 = hombre transgénero, a_5 = mujer transexual, a_6 = mujer transgénero y a_7 = no binaria},

se extrajo una muestra aleatoria de 81 participantes a = {4, 3, 2, 1, 2, 1, 1, 3, 3, 1, 2, 1, 2, 2, 1, 1, 2, 4, 2, 3, 4, 1, 2, 2, 2, 2, 1, 2, 2, 1, 1, 5, 6, 3, 2, 1, 1, 1, 5, 2, 2, 7, 2, 4, 4, 2, 1, 2, 2, 1, 2, 4, 6, 2, 3, 2, 5, 1, 1, 5, 2, 3, 7, 1, 1, 1, 2, 2, 2, 2, 2, 3, 2, 1, 7, 1, 2, 2, 1, 1, 2}. Se requiere representar su distribución por medio de una tabla de frecuencias y un diagrama de barras. Además, se desea calcular la moda de los 72 datos muestrales como medida de tendencia central, junto con la razón de variación universal (*RVU*) como medida de variación. Adicionalmente, se necesita obtener el intervalo de confianza al 95% para *RVU*. A tal fin se aconseja usar muestreo repetitivo con reemplazamiento (Bootstrap) por los métodos percentil (PERC) y percentil corregido de sesgo y acelerado (PCSA). Como información adicional, se necesita incluir la estimación puntual, error estándar y sesgo Bootstrap, así como la representación de la distribución Bootstrap mediante un histograma con la curva de densidad superpuesta, siguiendo la regla de Freedman y Diaconis (1981) para determinar la amplitud y el número de los intervalos de clase. Muestre los resultados redondeados a tres decimales.

```
# Definición del vector de datos
A <- c("Mujer", "Hombre", "Hombre transexual", "Hombre transgénero", "Mujer transexual", "Mujer transgénero", "No binario")
AA <- c("Mujer", "Hombre", "H. transex", "H. transg", "M. transex", "M. transg", "No binario")
a <- c(4, 3, 2, 1, 2, 1, 1, 3, 3, 1, 2, 1, 2, 2, 1, 1, 2, 4, 2, 3, 4, 1, 2, 2, 2, 2, 1, 2, 2, 1, 1, 5, 6, 3, 2, 1, 1, 1, 5, 2, 2, 7, 2, 4, 4, 2, 1, 2, 2, 1, 2, 4, 6, 2, 3, 2, 5, 1, 1, 5, 2, 3, 7, 1, 1, 1, 2, 2, 2, 2, 2, 3, 2, 1, 7, 1, 2, 2, 1, 1, 2)

# Creación de la tabla de frecuencias
tabla <- table(factor(a, levels = 1:7, labels = A))
identidad_sexual <- names(tabla)
frecuencia_absoluta <- as.vector(tabla)
N <- sum(frecuencia_absoluta)
frec_rel <- frecuencia_absoluta / N
porcentaje <- sprintf("%.1f%%", frec_rel * 100)
```

```
tabla_completa <- data.frame(identidad_sexual, frecuencia_absoluta, frecuencia_relativa = round(frec_rel, 3), porcentaje)
tabla_de_impresion <- tabla_completa[, c("identidad_sexual", "frecuencia_absoluta", "frecuencia_relativa", "porcentaje")]
cat("Distribución de frecuencias de la identidad sexual entre activistas de ONGs pro LGBT\n")
print(tabla_de_impresion)

# Diagrama de barras utilizando el programa ggplot2 y guardado como archivo JPEG
library(ggplot2)
marco_datos <- data.frame(Identidad_Sexual = factor(AA, levels = AA), Relative_Frequency = frec_rel)
diagrama <- ggplot(marco_datos, aes(x = Identidad_Sexual, y = Relative_Frequency)) +
geom_bar(stat = "identity", fill = "lightgreen", color = "black") +
labs(x = "Identidad sexual", y = "Frecuencia relativa") +
theme(axis.text.x.bottom = element_text(angle = 25, hjust = 0.5, size = 8), axis.text.y = element_text(size = 8), axis.title.x = element_text(size = 9), axis.title.y = element_text(size = 9), panel.background = element_rect(fill = "white"), panel.grid.major = element_blank(), panel.grid.minor = element_blank(), axis.line = element_line(color = "black"))
jpeg("diagrama1_de_barras.jpeg", width = 800, height = 600, units = "px", res = 300)
print(diagrama)
dev.off()
diagrama

# Cálculo de la moda
modas<-identidad_sexual[frecuencia_absoluta==max(frecuencia_absoluta)]
frecuencia_moda <- max(frec_rel)
n <- length(a)
k <- length(identidad_sexual)
c <- length(modas)
```

```
cat("Tamaño muestral: n =", n, "\n")
cat("Número de categorias nominales: k =", k, "\n")
cat("Categorías modales: mo =", modas, "\n")
cat("Número de valores modales: c =", c, "\n")
cat("Frecuencia relativa de la moda: fmo =", round(frecuencia_moda, 3), "\n")

# Cálculo de la Razón de Variación Universal (RVU) y su distribución en el muestreo Bootstrap
RVU <- k^2 /( k^2-1) * (1 - frecuencia_moda / c)
cat("Razón Universal de Variación: RVU =", round(RVU, 3), "\n")
set.seed(123)
B <- 1000
RVU_boot <- numeric(B)
for (i in 1:B) {
muestra_boot <- sample(a, replace = TRUE)
tabla_frecuencias_boot <- table(factor(muestra_boot, levels = 1:7, labels = A))
frecuencia_max_boot<-max(tabla_frecuencias_boot)/sum(tabla_frecuencias_boot)
num_frec_max_boot<-sum(tabla_frecuencias_boot==max(tabla_frecuencias_boot))
RVU_boot[i] <- k^2 / (k^2 - 1) * (1 - frecuencia_max_boot / num_frec_max_boot)
}
EB_RVU <- mean(RVU_boot)
sesgo_RVU <- mean(RVU_boot) - RVU
ee_RVU <- sd(RVU_boot)
cat("Estimación Bootstrap de RVU:", round(EB_RVU, 3), "\n")
cat("Sesgo Bootstrap de RVU:", round(sesgo_RVU, 3), "\n")
cat("Error estándar Bootstrap de RVU:", round(ee_RVU, 3), "\n")

#Histograma con la curva de densidad superpuesta (guardado como archivo JPEG)
datos_hist <- data.frame(RVU_boot)
```

```
    Cuartil1 <- quantile(datos_hist$RVU_boot, 0.25, type = 8)
    Cuartil3 <- quantile(datos_hist$RVU_boot, 0.75, type = 8)
    RIC <- Cuartil3 - Cuartil1
    FD <- 2 * RIC / (length(datos_hist$RVU_boot)^(1/3))
    histograma_boot <- ggplot(datos_hist, aes(x = RVU_boot)) +
    geom_histogram(binwidth = FD, fill = “lightblue”, color = “black”, aes(y = ..density..)) + geom_density(color = “black”, size = 1.5) + labs(x = “Distribución en el muestreo Bootstrap de RVU”, y = “Densidad”) + theme(panel.background = element_rect(fill = “white”), axis.text.x.bottom = element_text(size = 8), axis.text.y = element_text(size = 8), axis.title.x = element_text(size = 8), axis.title.y = element_text(size = 8), axis.line = element_line(color = “black”))
    jpeg(“histograma1_RVU.jpeg”, width = 800, height = 600, units = “px”, res = 300)
    print(histograma_boot)
    dev.off()
    histograma_boot

    # Intervalo de confianza Bootstrap al 95% utilizando el método percentil (PERC)
    IC_PERC_RVU <- quantile(RVU_boot, c(0.025, 0.975), type = 8)
    cat(«Intervalo de confianza Bootstrap al 95% por el método percentil para RVU: [«, round(IC_PERC_RVU[1], 3), «,», round(IC_PERC_RVU[2], 3), «]\n»)

    # Percentil corregido de sesgo de RVU
    z_0_RVU <- qnorm(sum(RVU_boot <= RVU) / B)
    if (is.infinite(z_0_RVU)) {z_0_RVU <- 0} else {z_0_RVU <- z_0_RVU}
    cat(“Percentil corregido de sesgo de RVU:”, round(z_0_RVU, 3), “\n”)

    # Factor de corrección de asimetría (aceleración) usando estimación Jackknife
    RVU_jackknife <- numeric(n)
    for (i in 1:n) {
```

```
    muestra_jackknife <- a[-i]
    tabla_frecuencia_jackknife<-table(factor(muestra_jackknife,levels
= 1:7, labels = A))
    frequencia_max_jackknife <- max(tabla_frecuencia_jackknife) /
sum(tabla_frecuencia_jackknife)
    num_max_frec_max_jackknife <- sum(tabla_frecuencia_jackk-
nife == max(tabla_frecuencia_jackknife))
    RVU_jackknife[i] <- k^2 / (k^2 - 1) * (1 - frequencia_max_jackk-
nife / num_max_frec_max_jackknife)
    }
    theta_jackknife <- sum(RVU_jackknife) / n
    acel<-sum((theta_jackknife-RVU_jackknife)^3)/(6*sum((theta_
jackknife - RVU_jackknife)^2)^(3/2))
    cat("Factor de corrección de asimetría (aceleración):", round(acel,
4), "\n")

    # Cálculo de los valores críticos o cuantiles de una distribución
normal estándar N(0, 1)
    z_LI <- qnorm(0.025)
    z_LS <- qnorm(0.975)

    # Orden de los cuantiles corregidos por sesgo y acelerados
(PCSA) para RVU
    LI_PCSA_RVU <- pnorm(z_0_RVU + (z_0_RVU + z_LI) / (1 -
acel * (z_0_RVU + z_LI)))
    LS_PCSA_RVU <- pnorm(z_0_RVU + (z_0_RVU + z_LS) / (1 -
acel * (z_0_RVU + z_LS)))

    # Intervalo de confianza Bootstrap por el método de percentil
corregido de sesgo y acelerado (PCSA) para RVU
    IC_PCSA_RVU <- quantile(RVU_boot, probs = c(LI_PCSA_
RVU, LS_PCSA_RVU), type = 8)
    cat("Intervalo de confianza Bootstrap al 95% por el método PCSA
para RVU: [", round(IC_PCSA_RVU[1], 3), ",", round(IC_PCSA_
RVU[2], 3), "]\n")
```

Este guion de instrucciones se puede ejecutar en línea al estar el programa R disponible en https://rdrr.io/snippets/ con más de 19,000 paquetes preinstalados con acceso gratuito, incluido el paquete ggplot2. No obstante, los gráficos que se obtienen son de baja definición. Para conseguir los dos archivos con gráficos de alta definición se requiere instalar el programa R ((R Foundation, 2024)) o RStudio (RStudio Team, 2023) en la computadora personal (de escritorio o portátil), descargar el paquete ggplot2 (desde la función paquetes → instalar paquete(s) de la barra de herramienta de R), instalarlo (install.packages(ggplot2)) y ejecutar el guion en la computadora.

A continuación, se muestran los resultados que arroja el guion de instrucciones desarrollado para el programa R. Estos incluyen una distribución de frecuencias (Tabla 1) y un diagrama de barras (Figura 1) de la distribución de la variable A en la muestra original de 81 participantes, así como el histograma de la distribución en el muestreo Bootstrap de *RVU* (Figura 2).

Tabla 1

Distribución de frecuencias de la identidad sexual entre activistas de ONGs pro LGBT

Identidad sexual	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Mujer	25	0.309	30.9%
Hombre	33	0.407	40.7%
Hombre transexual	8	0.099	9.9%
Hombre transgénero	6	0.074	7.4%
Mujer transexual	4	0.049	4.9%
Mujer transgénero	2	0.025	2.5%
No binario	3	0.037	3.7%

Tamaño muestral: $n = 81$
Numero de categorías nominales: $k = 7$
Categorías modales: *mo* = Hombre
Numero de valores modales: $c = 1$
Frecuencia relativa de la moda: $fmo = 0.407$
Razón Universal de Variación: $RVU = 0.605$

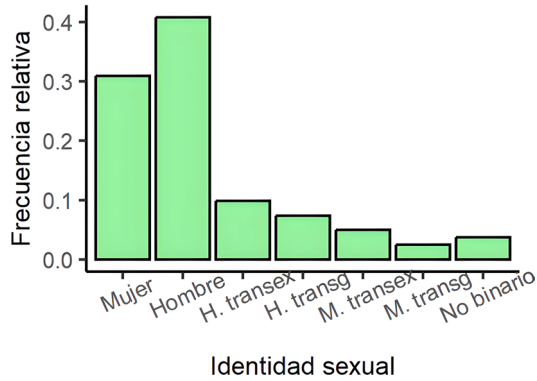

Figura 1. Diagrama de barras de la identidad sexual entre activistas de ONGs pro LGBT

Estimación Bootstrap de *RVU*: 0.603
Sesgo Bootstrap de *RVU*: -0.002
Error estándar Bootstrap de *RVU*: 0.06

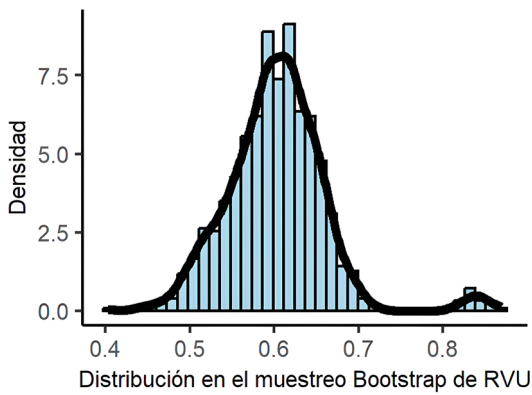

Figura 2. Histograma de la distribución en el muestreo de la Razón de Variación Universal

Intervalo de confianza Bootstrap al 95% por el método percentil para *RVU*: [0.492, 0.706]
Percentil corregido de sesgo de *RVU*: 0.166
Factor de corrección de asimetría (aceleración): -0.007
Intervalo de confianza Bootstrap al 95% por el método PCSA para *RVU*: [0.504, 0.838]

Ejemplo 2 de una muestra aleatoria con dos modas

De la variable cualitativa D = "preferencia sexual entre activistas de organizaciones no gubernamentales para la defensa de los derechos de las minorías sexuales" = {d_1 = asexual, d_2 = bisexual, d_3 = fluida, d_4 = heterosexual, d_5 = homosexual, d_6 = pansexual}, se extrajo una muestra aleatoria: b = {2, 5, 3, 2, 3, 5, 6, 5, 4, 2, 5, 4, 2, 3, 5, 2, 4, 4, 2, 3, 4, 6, 4, 5, 2, 2, 3, 4, 4, 3, 4, 6, 4, 5, 4, 4, 5, 5, 5, 5, 2, 5, 4, 3, 5, 2, 6, 2, 4, 3, 5, 4, 5, 2, 4, 4, 3, 5, 4, 5, 5, 4, 2, 5, 2, 2, 4, 5, 5, 1, 5, 4, 4, 4, 5, 4, 5, 4, 4, 5, 5}. Se requiere representar estos datos con una tabla de frecuencias y un diagrama de barras, así como calcular la moda como medida de tendencia central y la razón de variación universal (*RVU*) como medida de variación. Además, se desea obtener el intervalo de confianza al 95% por muestreo repetitivo con reemplazamiento por el método percentil (PERC) y percentil corregido de sesgo y acelerado (PCSA). Adicionalmente, se necesita la información sobre la estimación puntual, error estándar y sesgo Bootstrap, así como la representación de la distribución Bootstrap por medio de un histograma de densidad, determinando la amplitud y número de intervalos de clase por la regla de Freedman y Diaconis (1981). Extraer los resultados con 4 decimales.

```
# Definición del vector de datos muestrales
D <- c("Asexual", "Bisexual", "Fluida", "Heterosexual", "Homosexual",
"Pansexual")
d <- c(2, 5, 3, 2, 3, 5, 6, 5, 4, 2, 5, 4, 2, 3, 5, 2, 4, 4, 2, 3, 4, 6, 4, 5,
2, 2, 3, 4, 4, 3, 4, 6, 4, 5, 4, 4, 5, 5, 5, 5, 2, 5, 4, 3, 5, 2, 6, 2, 4, 3, 5, 4, 5,
2, 4, 4, 3, 5, 4, 5, 5, 4, 2, 5, 2, 2, 4, 5, 5, 1, 5, 4, 4, 4, 5, 4, 5, 4, 4, 5, 5)

# Creación de la tabla de frecuencias
tabla <- table(factor(d, levels = 1:6, labels = D))
orientacion_sexual <- names(tabla)
frecuencia_absoluta <- as.vector(tabla)
N <- sum(frecuencia_absoluta)
frec_rel <- frecuencia_absoluta / N
porcentaje <- sprintf("%.2f%%", frec_rel * 100)
```

```
tabla_completa <- data.frame(orientacion_sexual, frecuencia_
absoluta, frecuencia_relativa = round(frec_rel, 4), porcentaje)
tabla_de_impresion <- tabla_completa[, c("orientacion_sexual",
"frecuencia_absoluta", "frecuencia_relativa", "porcentaje")]
cat("Distribución de frecuencias de la orientación sexual entre
activistas de ONGs pro LGBT\n")
print(tabla_de_impresion)

# Diagrama de barras utilizando el programa ggplot2 y guardado
como archivo JPEG
library(ggplot2)
marco_datos <- data.frame(Orientacion_Sexual = factor(D,
levels = D), Relative_Frequency = frec_rel)
diagrama <- ggplot(marco_datos, aes(x = Orientacion_Sexual, y
= Relative_Frequency)) +
geom_bar(stat = "identity", fill = "lightyellow", color = "black") +
labs(x = "Orientacion sexual", y = "Frecuencia relativa") +
theme(axis.text.x.bottom = element_text(angle = 25, hjust = 0.5,
size = 8), axis.text.y = element_text(size = 8), axis.title.x = element_
text(size = 9), axis.title.y = element_text(size = 9), panel.background =
element_rect(fill = "white"), panel.grid.major = element_blank(), panel.
grid.minor = element_blank(), axis.line = element_line(color = "black"))
jpeg("diagrama2_de_barras.jpeg", width = 800, height = 600,
units = "px", res = 300)
print(diagrama)
dev.off()
diagrama

# Cálculo de la moda
modas<-orientacion_sexual[frecuencia_absoluta==max(frecuencia
_absoluta)]
frecuencia_moda <- max(frec_rel)
n <- length(d)
k <- length(orientacion_sexual)
c <- length(modas)
```

```
cat("Tamaño muestral: n =", n, "\n")
cat("Número de categorias nominales: k =", k, "\n")
cat("Categorías modales: mo =", modas, "\n")
cat("Número de valores modales: c =", c, "\n")
cat("Frecuencia relativa de la moda: fmo =", round(frecuencia_
moda, 4), "\n")

# Cálculo de la Razón de Variación Universal (RVU) y su distribu-
ción en el muestreo Bootstrap
RVU <- k^2 /( k^2-1) * (1 - frecuencia_moda / c)
cat("Razón Universal de Variación: RVU =", round(RVU, 4), "\n")
set.seed(123)
B <- 1000
RVU_boot <- numeric(B)
for (i in 1:B) {
muestra_boot <- sample(d, replace = TRUE)
tabla_frecuencias_boot <- table(factor(muestra_boot, levels = 1:6,
labels = D))
frecuencia_max_boot<-max(tabla_frecuencias_boot)/sum(tabla_
frecuencias_boot)
num_frec_max_boot<-sum(tabla_frecuencias_boot==max(tabla_
frecuencias_boot))
RVU_boot[i] <- k^2 / (k^2 - 1) * (1 - frecuencia_max_boot / num_
frec_max_boot)
}
EB_RVU <- mean(RVU_boot)
sesgo_RVU <- mean(RVU_boot) - RVU
ee_RVU <- sd(RVU_boot)
cat("Estimación Bootstrap de RVU:", round(EB_RVU, 4), "\n")
cat("Sesgo Bootstrap de RVU:", round(sesgo_RVU, 4), "\n")
cat("Error estándar Bootstrap de RVU:", round(ee_RVU, 4), "\n")

#Histograma con la curva de densidad superpuesta (guardado
como archivo JPEG)
datos_hist <- data.frame(RVU_boot)
```

```
Cuartil1 <- quantile(datos_hist$RVU_boot, 0.25, type = 8)
Cuartil3 <- quantile(datos_hist$RVU_boot, 0.75, type = 8)
RIC <- Cuartil3 - Cuartil1
FD <- 2 * RIC / (length(datos_hist$RVU_boot)^(1/3))
histograma_boot <- ggplot(datos_hist, aes(x = RVU_boot)) +
geom_histogram(binwidth = FD, fill = “darkolivegreen2”, color = “black”, aes(y = ..density..)) + geom_density(color = “black”, size = 1.5) + labs(x = “Distribución en el muestreo Bootstrap de RVU”, y = “Densidad”) + theme(panel.background = element_rect(fill = “white”), axis.text.x.bottom = element_text(size = 8), axis.text.y = element_text(size = 8), axis.title.x = element_text(size = 8), axis.title.y = element_text(size = 8), axis.line = element_line(color = “black”))
jpeg(“histograma2_RVU.jpeg”, width = 800, height = 600, units = “px”, res = 300)
print(histograma_boot)
dev.off()
histograma_boot

# Intervalo de confianza Bootstrap al 95% utilizando el método percentil (PERC)
IC_PERC_RVU <- quantile(RVU_boot, c(0.025, 0.975), type = 8)
cat(«Intervalo de confianza Bootstrap al 95% por el método percentil para RVU: [«, round(IC_PERC_RVU[1], 3), «,», round(IC_PERC_RVU[2], 4), «]\n»)

# Percentil corregido de sesgo de RVU
z_0_RVU <- qnorm(sum(RVU_boot <= RVU) / B)
if (is.infinite(z_0_RVU)) {z_0_RVU <- 0} else {z_0_RVU <- z_0_RVU}
cat(“Percentil corregido de sesgo de RVU:”, round(z_0_RVU, 4), “\n”)

# Factor de corrección de asimetría (aceleración) usando estimación Jackknife
RVU_jackknife <- numeric(n)
for (i in 1:n) {
```

```
    muestra_jackknife <- d[-i]
    tabla_frecuencia_jackknife <- table(factor(muestra_jackknife, levels = 1:6, labels = D))
    frequencia_max_jackknife <- max(tabla_frecuencia_jackknife) / sum(tabla_frecuencia_jackknife)
    num_max_frec_max_jackknife <- sum(tabla_frecuencia_jackknife == max(tabla_frecuencia_jackknife))
    RVU_jackknife[i] <- k^2 / (k^2 - 1) * (1 - frequencia_max_jackknife / num_max_frec_max_jackknife)
    }
    theta_jackknife <- sum(RVU_jackknife) / n
    acel<-sum((theta_jackknife-RVU_jackknife)^3)/(6*sum((theta_jackknife - RVU_jackknife)^2)^(3/2))
    cat("Factor de corrección de asimetría (aceleración):", round(acel, 4), "\n")

    # Cálculo de valores críticos o cuantiles de una distribución normal estándar N(0, 1)
    z_LI <- qnorm(0.025)
    z_LS <- qnorm(0.975)

    # Orden de los cuantiles corregidos por sesgo y acelerados (PCSA) para RVU
    LI_PCSA_RVU <- pnorm(z_0_RVU + (z_0_RVU + z_LI) / (1 - acel * (z_0_RVU + z_LI)))
    LS_PCSA_RVU <- pnorm(z_0_RVU + (z_0_RVU + z_LS) / (1 - acel * (z_0_RVU + z_LS)))

    # Intervalo de confianza Bootstrap por el método de percentil corregido de sesgo y acelerado (PCSA) para RVU
    IC_PCSA_RVU <- quantile(RVU_boot, probs = c(LI_PCSA_RVU, LS_PCSA_RVU), type = 8)
    cat("Intervalo de confianza Bootstrap al 95% por el método PCSA para RVU: [", round(IC_PCSA_RVU[1], 3), ",", round(IC_PCSA_RVU[2], 4), "]\n")
```

Cuando se ejecuta el guion de instrucciones se obtiene la siguiente salida, que incluye la distribución de frecuencia (Tabla 2) y el diagrama de barras (Figura 3) de la variable D en la muestra original de 81 participantes y el histograma de la distribución en el muestreo de *RVU* (Figura 4).

Tabla 2

Distribución de frecuencias de la orientación sexual entre activistas de ONGs pro LGBT

Orientación sexual	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Asexual	1	0.0123	1.23%
Bisexual	15	0.1852	18.52%
Fluida	9	0.1111	11.11%
Heterosexual	26	0.3210	32.10%
Homosexual	26	0.3210	32.10%
Pansexual	4	0.0494	4.94%

Tamaño muestral: n = 81

Numero de categorías nominales: k = 6
Categorías modales: *mo* = Heterosexual y Homosexual
Número de valores modales: c = 2
Frecuencia relativa de la moda: *fmo* = 0.321
Razón Universal de Variación: *RVU* = 0.8635

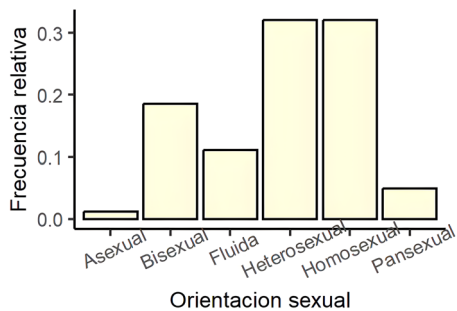

Figura 3. Diagrama de barras de la orientación sexual entre activistas de ONGs pro LGBT

Estimación Bootstrap de *RVU*: 0.6716
Sesgo Bootstrap de *RVU*: -0.1919
Error estándar Bootstrap de *RVU*: 0.0612

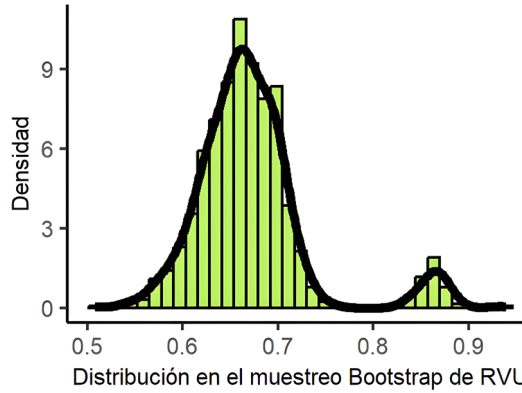

Figura 4. Histograma de la distribución en el muestreo de la Razón de Variación Universal

Intervalo de confianza Bootstrap al 95% por el método percentil para *RVU*: [0.5840, 0.8698]
Percentil corregido de sesgo de *RVU*: 1.9110
Factor de corrección de asimetría (aceleración): -0.0110
Intervalo de confianza Bootstrap al 95% por el método PCSA para *RVU*: [0.8630, 0.9397]

Conclusiones

Al acudir a muestreo repetitivo con reemplazamiento (Bootstrap), se puede obtener el intervalo de confianza e incluso el error estándar para la Razón de Variación Universal (*RVU*). Este último es la desviación estándar corregida de sesgo de la distribución en el muestreo Bootstrap del estimador o estadístico. Cuando la distribución en el muestreo del estadístico es desconocida, como en el caso de *RVU*, la mejor opción es un método Bootstrap no paramétrico. Entre estos destacan el método percentil, PERC, y el percentil corregido de sesgo y acelerado, PCSA (Rousselet et al., 2021). Cuando el sesgo y la aceleración (asimetría) son pequeños (|*sesgo*| o |*a*| < 0.1), ambos métodos son perfectamente válidos, como ocurre en el primer ejemplo, donde el sesgo Bootstrap de *RVU* es de -0.002 y el factor de corrección de asimetría o aceleración es de -0.007. No obstante, cuando al menos uno de estos dos índices

es moderado (|*sesgo*| o |*a*| ≥ 0.1) es mejor el método de percentil corregido de sesgo y acelerado (Efron & Narasimhan, 2020), como se observa en el segundo ejemplo, donde el sesgo Bootstrap de *RVU* es moderado, con un valor de -0.1919, aunque el factor de corrección de asimetría o aceleración es bajo, con un valor de -0.0110.

El mayor problema del método de percentil corregido de sesgo y acelerado surge cuando el percentil corregido de sesgo (z_0) presenta un valor extremo. En este caso, se puede dar a z_0 un valor nulo, lo que arroja un resultado equivalente al método percentil, como Efron (1987) sugiere. En ninguno de los dos ejemplos mostrados se dio esta situación. En caso de presentarse, aparecería como resultado: "Percentil corregido de sesgo de *RVU*: 0" y se observaría que los límites del intervalo de confianza PERC y PCSA son muy parecidos.

El procedimiento de remuestreo con reemplazamiento permite observar que la distribución en el muestreo de *RVU* tiene el perfil bimodal de una distribución continua, con dos modas asimétricas, una moda mayor a la izquierda y una moda menor a la derecha, con una agrupación de datos bastante simétrica alrededor de estas dos modas y un valle o depresión entre ambas. Este perfil se aleja claramente de la forma acampanada y simétrica en torno a una única moda de la distribución normal.

Cabe observar que, en el guion de instrucciones desarrollado en este artículo, se usa una semilla para la generación de muestras. Este procedimiento se implementa para que la estimación sea más estable o reproducible. La elección de la semilla (123) es arbitraria y puede ser cualquier número. Sin embargo, es común utilizar números fijos o patrones simples para facilitar la reproducibilidad y la comunicación del código (Canty, 2022). Por otra parte, la estimación de la asimetría (aceleración) se realiza por el método de la navaja (Jackknife), lo que hace que el resultado sea totalmente estable.

Es importante remarcar que este guion de instrucciones debe aplicarse con una muestra aleatoria, representativa de la población y grande, de por lo menos 30 datos muestrales, para que la estimación sea válida (Efron & Narasimhan, 2020).

Este guion es perfectamente aplicable a variables de categorías ordenadas, como podría ser la orientación sexual medida con la escala de Kinsey, que cuenta con siete categorías ordenadas (Zietsch & Sidari, 2020). La aplicación de *RVU* con variables cuantitativas discretas con muchos valores (por ejemplo, número de parejas sexuales del mismo sexo) y variables continuas (por ejemplo, volumen del flujo sanguíneo de la arteria peneana medido por un pletismógrafo ante una situación controlada de estímulos homoeróticos) es posible, sobre todo si su distribución tiene un pico definido. No obstante, se desaconseja por una infrautilización de la información contenida en los datos frente a las medidas absolutas o relativas basadas en el promedio o la mediana de puntuaciones diferenciales con respecto a la media aritmética o la mediana (Ramachandran & Tsokos, 2020).

Con variables cuantitativas discretas, la moda se sigue estimando por el método del valor de frecuencia máxima como con las variables cualitativas y ordinales (Coolidge, 2020). No obstante, con variables cuantitativas continuas, se requiere utilizar un método de estimación de la densidad y buscar el valor con densidad máxima, como el método desarrollado por Parzen (1962) o el de semimuestras reiterativas de Bickel (2002). A tal fin, se puede aplicar el programa R (Poncet, 2022). Un problema adicional con muestras de variables cuantitativas continuas es que se requiere discretizar la distribución en k intervalos de clase, lo que incrementa más la pérdida de información (Banić & Elezović, 2021). Para determinar el número de intervalos de clase, una buena opción es la regla de optimización sin asumir normalidad de Freedman y Diaconis (Contreras-Reyes & Brito, 2022).

Cuando se describen variables cualitativas, se sugiere calcular la moda como medida de tendencia central, así como la razón de variación universal de Moral-de la Rubia (2022a) y la desviación estándar desde la moda de Kvalseth (1988) como medidas de variación. Esta última usa más información de la distribución que las medidas de variación y se ve menos influenciada por el efecto techo ante la proximidad de la distribución limitante (uniforme) que el índice de variación cualitativa de Gibbs y Poston y la entropía relativa de Shannon

(Moral-de la Rubia, 2022a). A su vez, se invita a aplicar la asimetría y el apuntamiento de Moral-de la Rubia (2021, 2022b, 2023) como medidas de la forma, aparte de utilizarse la tabla de frecuencias y el diagrama de barras o circular. Asimismo, es importante seguir trabajando en el desarrollo de medidas descriptivas para estas variables, dado que es un tema escasamente abordado por la estadística (Li et al., 2020), cuando son variables muy importantes en la investigación en psicología y otras ciencias sociales y de la salud (Levitt, 2021; Maxwell, 2021).

Referencias

ATLAS.ti Scientific Software Development GmbH. (2023). *ATLAS.ti Windows (versión 23.2.1)* [Software de análisis de datos cualitativos]. https://atlasti.com

Banić, N., & N. Elezović, N. (2021). TVOR: Finding discrete total variation outliers among histograms. *IEEE Access, 9*, 1807-1832. https://doi.org/10.1109/ACCESS.2020.3047342

Bickel, D. R. (2002). Robust estimators of the mode and skewness of continuous data. *Computational Statistics and Data Analysis, 39*, 153-163. https://doi.org/10.1016/S0167-9473(01)00057-3

Canty, A. (2022). *Package 'boot'*. https://cran.r-project.org/web/packages/boot/boot.pdf

Contreras-Reyes, J. E., & Brito, A. (2022). Refined cross-sample entropy based on Freedman-Diaconis rule: application to foreign exchange time series. *Journal of Applied and Computational Mechanics, 8*(3), 1005-1013. https://doi.org/10.22055/JACM.2022.39470.3412

Coolidge, F. L. (2020). *Statistics: A gentle introduction* (4th ed.). Sage Publications. https://doi.org/10.4135/9781071939000

Deacon, D., & Stanyer, J. (2021). Media diversity and the analysis of qualitative variation. *Communication and the Public, 6*(1-4), 1932. https://doi.org/10.1177/20570473211006481

Efron, B. (1987) Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association, 82*(397), 171-185. https://doi.org/10.2307/2289144

Efron, B., & Narasimhan, B. (2020) The automatic construction of bootstrap confidence intervals. *Journal of Computational and Graphical Statistics, 29*(3), 608-619. https://doi.org/10.1080/10618600.2020.1714633

Evren, A., & Erhan, U. (2017). Measures of qualitative variation in the case of maximum entropy. *Entropy, 19*(5), 204. https://doi.org/10.3390/e19050204

Evren, A., Tuna, E., Ustaoglu, E., & Sahin, B. (2021). Some dominance indices to determine market concentration. *Journal of Applied Statistics, 48*(13-15), 2755-2775. https://doi.org/10.1080/02664763.2021.1963421

Freeman, L. C. (1965). *Elementary applied statistics for students in behavioral sciences*. John Wiley and Sons. https://doi.org/10.2307/3538646

Freedman, D., & Diaconis, P. (1981) On the histogram as a density estimator: L2 theory. *Probability Theory and Related Fields, 57*(4), 453-476. https://doi.org/10.1007/BF01025868

Gibbs, J. P., & Poston, D. L, Jr. (1975) The division of labor: conceptualization and related measures. *Social Forces, 53*(3), 468-476. https://doi.org/10.2307/2576589

Golan, A., & Harte, J. (2022) Information theory: a foundation for complexity science. *Proceedings of the National Academy of Sciences,* 119(33), e2119089119. https://doi.org/10.1073/pnas.2119089119

Goss-Sampson, M. A. (2020). *Statistical analysis in JASP: A guide for students* (4th ed.). University of Greenwich. https://doi.org/10.6084/m9.figshare.9980744

Hammack, P. L., Hughes, S. D., Atwood, J. M., Cohen, E. M., & Clark, R. C. (2022). Gender and sexual identity in adolescence: A mixed-methods study of labeling in diverse community set-

tings. *Journal of Adolescent Research, 37*(2), 167-220. https://doi.org/10.1177/07435584211000315

Hyndman, R. J., & Fan, Y. (1996) Sample quantiles in statistical packages. *American Statistician, 50*(4), 361-365. https://doi.org/10.2307/2684934

IBM Documentation. (2023). *IBM SPSS Statistics 29 Documentation*. https://www.ibm.com/support/pages/ibm-spss-statistics-29-documentation

Kvalseth, T. O. (1988). Measuring Variation for Nominal Data. *Bulletin of the Psychonomic Society, 26*(5), 433-436. https://doi.org/10.3758/BF03334906

Levitt, H. M. (2021). Qualitative generalization, not to the population but to the phenomenon: reconceptualizing variation in qualitative research. *Qualitative Psychology, 8*(1), 95-110. https://doi.org/10.1037/qup0000184

Li, Y., Garg, H., & Deng, Y. (2020). A New Uncertainty Measure of Discrete Z-Numbers. *International Journal of Fuzzy Systems, 22*, 760-776. https://doi.org/10.1007/s40815-020-00819-8

Maxwell, J. A. (2021) Why qualitative methods are necessary for generalization. *Qualitative Psychology, 8*(1), 111-118. https://doi.org/10.1037/qup0000173

Microsoft. (2021). Microsoft Excel 2021. Microsoft Corp.

Moral-de la Rubia, J. (2021). Una medida de asimetría unidimensional para variables cualitativas. *Revista de Psicología, 41*(1), 421-459. https://doi.org/10.18800/psico.202201.017

Moral-de la Rubia, J. (2022a). Una medida de variación para datos cualitativos con cualquier tipo de distribución. *Psychologia, 16*(2), 63-76. https://doi.org/10.21500/19002386.5642

Moral-de la Rubia, J. (2022b). Medición del apuntamiento en variables en escala nominal. *Revista de Psicología, 41*(1), 421-459. https://doi.org/10.18800/psicologia.202301.016

Moral-de la Rubia, J. (2023). Shape measures for the distribution of a qualitative variable. *Open Journal of Statistics, 13*(4), 619-634. https://doi.org/10.4236/ojs.2023.134030

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics, 33*(3), 10651076. https://doi.org/10.1214/aoms/1177704472

Poncet, P. (2022) *Package 'modeest'*. Mode estimation. https://cran.r-project.org/web/packages/modeest/modeest.pdf

R Foundation. (2024). *The R project for statistical computing*. https://www.r-project.org/

Ramachandran, K. M., & Tsokos, C. P. (2020). *Mathematical statistics with applications in R*. Academic Press.

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2021) The Percentile Bootstrap: A Primer with Step-by-Step Instructions in R. *Advances in Methods and Practices in Psycho-logical Science, 4*(1), https://doi.org/10.1177/2515245920911881

Rousselet, G., Pernet, C. R., & Wilcox, R. R. (2023). An introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations. *Meta-Psychology, 7*. https://doi.org/10.15626/MP.2019.2058

RStudio Team. (2023). *RStudio: integrated development for R (version 2023.06.0-421)* [Computer software]. RStudio, PBC. http://www.rstudio.com/

Shannon, C. E. (1948). A Mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Simpson, E. H. (1949). Measurement of diversity. *Nature, 163*, 688-688. http://dx.doi.org/10.1038/163688a0

StataCorp. (2023). *Stata 18. Getting started with Stata for Windows*. Stata Press.

VERBI Software. (2024). *MAXQDA 2024* [computer software]. VERBI Software. https://www.maxqda.com/

Weiss, C. H. (2019). On the sample coefficient of nominal variation. A. Steland, E. Rafajłowicz & O. Okhrin (Eds.), *Stochastic models, statistics and their applications* (vol. 294, pp. 239-250). Springer Proceedings in Mathematics & Statistics. https://doi.org/10.1007/978-3-030-28665-1_18

Wilcox, A. R. (1973). Indices of qualitative variation and political measurement. *The Western Political Quarterly, 26*(2), 325-343. https://doi.org/10.2307/446831

Zaiontz, C. (2024). *Real Statistics using Excel*. https://real-statistics.com/

Zelikman, E., Wu, Y., Mu, J., & Goodman, N. (2022). Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems, 35*, 15476-15488.

Zietsch, B. P., & Sidari, M. J. (2020). The Kinsey scale is ill-suited to most sexuality research because it does not measure a single construct. *Proceedings of the National Academy of Sciences, 117*(44), 27080-27080. https://doi.org/10.1073/pnas.2015820117

Recibido: 19/03/2024
Revisado: 25/02/2025
Aceptado: 27/03/2025