

ARTÍCULO DE INVESTIGACIÓN

DOS MEDIDAS DE ASIMETRÍA BASADAS EN LA MODA:
CÁLCULO Y NORMAS INTERPRETATIVAS

TWO MEASURES OF SKEWNESS BASED ON MODE: CALCULATION AND
INTERPRETATIVE RULES

JOSÉ MORAL DE LA RUBIA¹

FECHA DE RECEPCIÓN 08/06/2022 • FECHA DE ACEPTACIÓN 23/06/2023

Para citar este artículo: Moral de la Rubia, J. (2023). Dos medidas de asimetría basadas en la moda: cálculo y normas interpretativas, *Psychologia. Avances de la Disciplina*, 17(2) 39-54. <https://doi.org/10.21500/19002386.6542>

Resumen

Este estudio metodológico de simulación presenta de forma ejemplificada dos medidas de asimetría. Aunque pueden ser útiles cuando la distribución es unimodal, no se reportan en la investigación psicológica. Una es la distancia estandarizada de la media a la moda de Pearson. La otra es la medida robusta de asimetría de Bickel. Se muestra cómo calcular la estimación puntual y de intervalo con el programa R. Además, se calculan intervalos de confianza al 90 %, 95 % y 99 % con 10 000 extracciones con reemplazamiento de muestras-población con distribución normal y diferentes tamaños para disponer de directrices interpretativas de simetría. Se concluye que la regla ∓ 0.1 no aplica, la moda de Grenander proporciona los intervalos de confianza más eficientes, pero la asimetría de Bickel es la opción con variables ordinales.

Palabras clave: simetría, estimación de la moda, bootstrap, descriptivos, inferencia estadística.

Abstract

This methodological study of simulation presents in exemplified form two measures of skewness. Although they may be useful when the distribution is unimodal, they are not reported in psychological research. One is Pearson's standardized distance from the mean to the mode. The other is the Bickel's robust measure of skewness. It is shown how to compute the point and interval estimate with the R program. Moreover, interval confidences at 90%, 95% and

¹ Facultad de Psicología, Universidad Autónoma de Nuevo León, c/ Dr. Carlos Canseco 110, Col. Mitras Centro, CP 64460, Monterrey, NL, México. Tel. 00 52 81 8333 82 33. Dirección electrónica: jose.morald@uanl.edu.mx. ORCID: <http://orcid.org/0000-0003-1856-1458>

99% are calculated with 10 000 draws with replacement from normally distributed samples-population with different sizes to have interpretative guidelines for symmetry. It is concluded that the ∓ 0.1 rule does not apply with these measures, Grenander's mode provides the most efficient confidence intervals, but Bickel's skewness is the option with ordinal variables.

Key words: symmetry, estimation of mode, bootstrap, descriptive measures, statistical inference.

Introducción

El primer objetivo de este estudio metodológico es mostrar la estimación puntual y por intervalo de la medida de asimetría basada en la distancia estandarizada de la media a la moda de Pearson (1894), que es para variables cuantitativas, y la medida robusta de asimetría de Bickel (2002), que es para variables tanto cuantitativas como ordinales, usando el programa R. Para la medida de Pearson, se utilizan dos estimadores de la moda: la función de densidad de probabilidad empírica de Parzen (1962) y el estimador de Grenander (1965). Para la medida de Bickel, la moda se estima mediante su método de semimuestras reiterativas con variables cuantitativas y el método del valor de frecuencia máxima con variables ordinales. El segundo objetivo es calcular intervalos de confianza al 90 %, 95 % y 99 % con 10 000 extracciones aleatorias con reemplazamiento de muestras-población con distribución normal y diferentes tamaños para disponer de directrices interpretativas de estas dos medidas de asimetría. La distribución normal, que es simétrica y de colas medias, se adopta como referencia de simetría nula por su centralidad en la teoría de la probabilidad y por ser un modelo probabilístico seguido por variables cuantitativas en diversos campos científicos. Desde las normas generadas por simulación, se revisa la regla interpretativa del 0 ∓ 0.1 para simetría.

Concepto y medición de la asimetría

La *asimetría* es una propiedad de la forma de la distribución empírica. Definido un eje de simetría para dividir la distribución en dos partes, como la media aritmética, hay simetría si ambas partes de la distribución son iguales; por el contrario, hay asimetría si ambas partes son dispares (Moral, 2023; Shi et al. 2020).

Las *medidas de asimetría relativas* o libres de unidad de medida se definen como cocientes, proporciones o promedios centrados en 0 (Khan, Cheema, Hussain & Abdel-Salam, 2021). El valor de 0 indica asimetría nula, es decir,

simetría. En una distribución unimodal continua, refleja que los dos hombros y las dos colas a ambos lados del eje de simetría son idénticos, esto es, un lado es el reflejo del otro. En este tipo de distribución, un valor positivo evidencia que la cola derecha es más larga que la izquierda, y, por el contrario, un valor negativo revela que la cola izquierda es más larga que la derecha (Sarka, 2021).

Se pueden distinguir cuatro tipos de medidas de asimetría (Eberl & Klar, 2019; Gupta & Kapoor, 2020). Un primer tipo se basa en el tercer momento central o tercer cumulante estandarizado; un segundo tipo, en los cuantiles o los expectiles; un tercer tipo son medidas mixtas que se basan en momentos y cuantiles, y, finalmente, existe un cuarto tipo de medidas robustas; una está basada en medias truncadas de forma simétrica (25 % a ambos lados) y asimétrica, 5 % a cada extremo (Bono, Arnau, Alarcón & Blanca, 2020) y otra en la estimación robusta de la moda (Bickel, 2002).

Distancia estandarizada de la media aritmética a la moda

La primera medida de asimetría de la distribución de una variable aleatoria fue la distancia estandarizada de la media aritmética a la moda y fue desarrollada por Pearson (1894). Se obtiene mediante un cociente. Su numerador es la diferencia entre la media aritmética (primer momento) y la moda que tiene que ser única. Su denominador es la desviación estándar (raíz cuadrada del segundo momento central). Se puede denotar por A_{p1} , al ser la primera fórmula desarrollada por Pearson (1894).

$$A_{p1} = \frac{\mu_x - Mo(X)}{\sigma_x}$$

μ_x = media aritmética o esperanza matemática poblacional de X. Si X es una variable continua con soporte $(-\infty, \infty)$ y función de densidad $f_x(x)$, su media es la integral indefinida del producto de la variable X y su función de densidad. Si X tiene un soporte acotado, la media es la integral definida en dicho soporte.

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Si X es una variable discreta con soporte $\{1, 2, \dots, n\}$ y función de masa de probabilidad $f_X(x) = P(X = x)$, su media aritmética es la suma del producto de los valores de X y sus correspondientes probabilidades. Lo mismo aplicaría para un soporte no finito que abarque toda la recta de los números enteros o naturales.

$$\mu_X = E(X) = \sum_{i=1}^n x_i P(X = x_i)$$

$Mo(X)$ = moda de X . Si X es una variable continua con soporte $(-\infty, \infty)$ y función de densidad $f_X(x)$, su moda corresponde al valor o los valores en los cuales su función de densidad alcanza su máximo o pico. Lo mismo aplicaría para un soporte acotado.

$$Mo(X) = \{x_k | \text{Max}[f_X(x_i)] = f_X(x_k)\}$$

Si X es una variable discreta con dominio $\{1, 2, \dots, n\}$ y función de masa de probabilidad $f_X(x) = P(X = x)$, su moda es el valor o los valores con la probabilidad máxima. Lo mismo aplicaría para un soporte no finito que abarque toda la recta de los números enteros o naturales.

$$Mo(X) = \{x_k | \text{Max}[P(X = x)] = P(x_k)\}$$

σ_X = desviación estándar de X . Si X es una variable continua con soporte $(-\infty, \infty)$ y función de densidad $f_X(x)$, su desviación estándar es la raíz cuadrada de la integral indefinida del producto del cuadrado de las puntuaciones diferenciales (con respecto a la media) y la función de densidad. Si X tiene un soporte acotado, la integral queda definida en dicho soporte.

$$\sigma_X = \sqrt{E[(X - \mu)^2]} = \sqrt{\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx}$$

Si X es una variable discreta con soporte $\{1, 2, \dots, n\}$ y función de masa de probabilidad $f_X(x) = P(X = x)$, la desviación estándar es la raíz cuadrada de la suma de productos del cuadrado de las puntuaciones diferenciales (con respecto a la media) y sus correspondientes probabilidades. Lo mismo aplicaría para un soporte no

finito que abarque toda la recta de los números enteros o naturales.

$$\sigma_X = \sqrt{E[(X - \mu_X)^2]} = \sqrt{\sum_{i=1}^n (x_i - \mu_X)^2 P(X = x_i)}$$

Cuando la asimetría se calcula en una muestra aleatoria de n datos de X , se usa la media muestral (\bar{x}), la moda muestral (mo) y la desviación estándar muestral o corregida de sesgo (s_{n-1}).

$$x = \{x_i\}_{i=1}^n = \{x_1, x_2, \dots, x_n\} \subseteq X$$

$$A_{P_1} = \frac{\bar{x} - mo}{s_{n-1}}$$

Media muestral:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Moda muestral: mo = valor muestral con mayor frecuencia

$$mo = \{x_k | \text{max}(n_{x_i}) = n_{x_k}\}$$

Desviación estándar muestral o corregida de sesgo:

$$s_{n-1} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Estimación de la moda desde el histograma

Cuando la variable es continua y aparecen múltiples modas en la muestra aleatoria, se puede representar los datos por medio de un histograma de frecuencias. Se requiere determinar el número de intervalos de clase (k) y su amplitud (a), de preferencia homogénea o constante. Esto se puede hacer mediante la regla de Scott (1979), en caso de distribución normal: $a = (3.49 \times s_{n-1})/n^{1/3}$, o la regla de Freedman y Diaconis (1981), en caso de no normalidad, $a = (2 \times R_{IC})/n^{1/3}$, donde s_{n-1} es la desviación estándar muestral o corregida de sesgo y R_{IC} es el rango intercuartílico. Ambas reglas determinan la amplitud y esta determina el número de intervalos: $k = [R/a] = [(\text{max}(x_i) - \text{min}(x_i))/a]$. Otra opción es la regla de la Universidad de Rice (Lane, 2015), la cual define el número de intervalos ($k = 2n^{1/3}$) y de ahí la amplitud ($a = R/k$). Se trata de una simplificación de las dos reglas previas (Moral, 2023). Si en el histograma se observa un intervalo modal, esto es, con altura máxima, el valor de

la moda se obtiene por interpolación lineal mediante una de las dos siguientes reglas:

Sea una muestra aleatoria de X integrada por n datos agrupados en k intervalos de clase con frecuencia absoluta n_i , donde $i = 1, 2, \dots, k$, y $\sum n_i = n$.

Intervalo modal:

$$[LI_c, LS_c]; \max(\{n_i\}_{i=1}^k) = n_c; 1 \leq c \leq k$$

Amplitud del intervalo homogéneo o constante: $a =$

$$LS_c - LI_c$$

$$mo = LI_c + \frac{n_{c+1}}{n_{c-1} + n_{c+1}} a$$

$$mo = LI_c + \frac{d_1}{d_1 + d_2} a = LI_c + \frac{n_c - n_{c-1}}{(n_c - n_{c-1}) + (n_c - n_{c+1})} a$$

Para los ejemplos, se genera con el programa R una muestra aleatoria de 16 datos correspondiente a una supuesta prueba de aptitud que tiene distribución normal de media $\mu = 5.5$ y desviación estándar $\sigma = 2$. Las puntuaciones se dan redondeadas a dos decimales.

$$x <- rnorm(16, 5.5, 2) \\ \text{round}(x, 2)$$

$x = \{5.26, 7.58, 1.43, 2.90, 6.37, 8.51, 6.31, 4.63, 4.92, 4.02, 4.09, 6.09, 3.83, 3.32, 6.62, 3.71\}$

Se define la amplitud de los intervalos por la regla de Scott (1979).

$$a = \frac{3.49 \times s_{n-1}}{\sqrt[3]{n}} = \frac{3.49 \times 1.851}{\sqrt[3]{16}} = 2.563;$$

$$k = \frac{R}{a} = \left\lceil \frac{8.51 - 1.43}{2.563} \right\rceil = \left\lceil \frac{7.08}{2.563} \right\rceil = 3$$

Resultan tres intervalos de clase con amplitud de 2.563. Sus frecuencias aparecen en la Tabla 1. La Figura 1 muestra el histograma en el que se aprecia un intervalo modal.

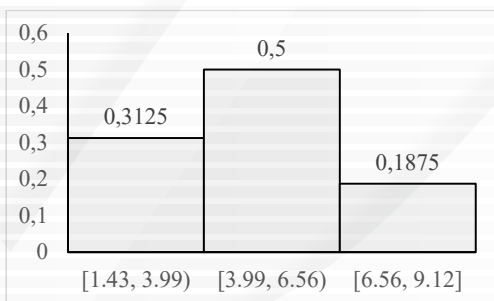


Figura 1. Histograma de frecuencias.

Tabla 1

Frecuencias por intervalos de clase

i	LI_i	LS_i	x_i	n_i	f_i
1	1.43	3.993	2.712	5	0.3125
2	3.993	6.557	5.275	8	0.5
3	6.557	9.120	7.838	3	0.1875
Σ				16	1

Nota. i = indicador de clase, LI_i = límite inferior, LS_i = límite superior, x_i = marca de clase, n_i = frecuencia absoluta simple, f_i = frecuencia relativa simple.

Intervalo modal o con frecuencia máxima: [3.993, 6.557).

$$mo = x_c + \frac{n_{c+1}}{n_{c-1} + n_{c+1}} a = 3.993 + \frac{3}{5 + 3} \times 2.563 = 4.955$$

$$mo = x_i + \frac{d_1}{d_1 + d_2} a = 3.993 + \frac{8 - 5}{(8 - 5) + (8 - 3)} \times 2.563 = 4.955$$

La moda coincide con ambas reglas de interpolación en este ejemplo y su valor es 4.955. El valor del coeficiente de asimetría de Pearson (1894) está muy próximo a cero, como se esperaba al proceder la muestra de una distribución normal.

$$A_{P1} = \frac{\bar{x} - mo}{s_{n-1}} = \frac{4.974 - 4.955}{1.851} = 0.010$$

La moda de Parzen

Otra opción para estimar la moda es usar el valor x con densidad máxima en la función de densidad de probabilidad empírica: $f_x(x)$, lo que se puede calcular mediante la librería de R, denominada *modest* (Poncet, 2022). Desde este planteamiento, se elige una de las funciones de Kernel y un ancho de banda para estimar la densidad (Parzen, 1962). El Kernel gaussiano y el ancho de banda recomendado por Silverman (1986) suelen ser la elección más común o práctica.

Sea una muestra aleatoria de n datos de la variable continua X .

$$x = \{x_i\}_{i=1}^n = \{x_1, x_2, \dots, x_n\}$$

Kernel gaussiano

$$\hat{f}(x \in x) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^n e^{-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2}$$

Ancho de banda recomendado por Silverman (1986)

$$h = \frac{0.9\hat{\sigma}}{\sqrt[5]{n}} = 0.9 \times \hat{\sigma} \times n^{-1/5}$$

$$R_{IQ}(Z) = z_{0.75} - z_{0.25} = 0.6745 - (-0.6745) \\ = 2 \times 0.6745 = 1.349; Z \sim N(0, 1)$$

$$\hat{\sigma} = \min(s_{n-1}(x), R_{IQ}(x)/R_{IQ}(Z)) \\ = \min(s_{n-1}(x), R_{IQ}(x)/1.349)$$

Desviación estándar muestral:

$$s_{n-1}(x) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Rango intercuartílico:

$$R_{IQ} = P_{75}(x) - P_{25}(x)$$

Para el cálculo de los cuantiles, el programa R usa por defecto la regla 7 o moda de la i -ésima estadística de orden de una distribución uniforme estándar, la cual sigue una distribución Beta($\alpha = i, \beta = n + 1$).

p = orden del cuantil

$$mo(i) = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{i - 1}{n - 1} = p, i \sim Beta(\alpha = i, \beta = n + 1)$$

i = orden del dato

$$i = 1 + p(n - 1)$$

Cuantil de orden p o valor de X en la i -ésima posición.

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$C_X(p) = x_{(i)} = x_{(li)} + (i - [i])(x_{(li+1)} - x_{(li)})$$

Moda estimada:

$$\widehat{mo} = \left\{ x_c \mid \max \left(\{ \hat{f}(x_i) \}_{i=1}^n \right) = \hat{f}(x_c) \right\}, \text{ donde } 1 \leq c \leq n$$

Si la moda de Parzen (1962) se calcula por el programa R con el kernel gaussiano y el ancho de banda de Silverman (1986), las instrucciones son las siguientes.

```
library(modeest)
x<-c(5.26, 7.58, 1.43, 2.90, 6.37, 8.51, 6.31, 4.63,
4.92, 4.02, 4.09, 6.09, 3.83, 3.32, 6.62, 3.71)
mo=mlv(x, method = "parzen", bw=bw.nrd0(x),
kernel = "gaussian")
cat(»Moda de Parzen = »,mo,«\n«)
```

El resultado es el siguiente: Moda de Parzen = 4.229

Cabe señalar que el programa R tiene otros métodos para definir la función de Kernel aparte del gaussiano o normal, como el uniforme (kernel = »uniform«), bponderado (»biweight«), coseno (»cosine«), »eddy«, »epanechnikov«, coseno optimizado (»optcosine«), »rectangular« y »triangular«. El más usado por sus propiedades matemáticas es el normal, pero el Kernel de Epanechnikov (1969) es el que minimiza el error cuadrático medio integrado y es el más recomendado. Se trata de una función parabólica y depende del ancho de banda. Se indica usar un ancho de banda que minimice el error cuadrático medio integrado asintótico y se recomienda el método de Sheather y Jones (1991) para estimar su valor (Guidoum, 2020; Henderson, Papadopoulos & Parmeter, 2023).

$$\hat{f}_h(x \in x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

$$K(u) = \frac{3}{4}(1 - u^2); |u| \leq 1$$

$$\hat{f}_h(x \in x) = \frac{3}{4nh} \sum_{i=x-h}^{x+h} \left[1 - \left(\frac{x - x_i}{h}\right)^2 \right]$$

Si la moda de Parzen (1962) se calcula por el programa R con el kernel de Epanechnikov (1969) y el ancho de banda de Sheather y Jones (1991), las instrucciones son las siguientes:

```
library(modeest)
x<-c(5.26, 7.58, 1.43, 2.90, 6.37, 8.51, 6.31, 4.63,
4.92, 4.02, 4.09, 6.09, 3.83, 3.32, 6.62, 3.71)
mo=mlv(x, method = »parzen«, bw=«SJ«, kernel
= »epanechnikov«)
cat(»Moda = »,mo,«\n«)
```

El resultado es el siguiente: Moda = 4.075.

No hay una regla de interpretación general para el coeficiente A_{pi} . No obstante, se pueden calcular el error e intervalos de confianza por el método de muestreo repetitivo con reposición (bootstrap) mediante el método de percentiles corregidos de sesgo y acelerados (bca, por sus siglas en inglés) o el normal. Se recomienda un nivel de significación del 90 % con muestras pequeñas y medias y del 95 % con muestras grandes. El programa R permite este cálculo (Canty & Ripley,

2022). El ancho de banda se define por la regla de Silverman (1986), ya que el procedimiento de Sheather y Jones (1991) genera usualmente errores que detienen el cálculo. Es muy importante complementar el análisis con una inspección visual de los datos a través del histograma y una gráfica de caja y bigotes, lo que se añade a las instrucciones de R.

```
library(modeest)
x<- c(5.26, 7.58, 1.43, 2.90, 6.37, 8.51, 6.31, 4.63,
4.92, 4.02, 4.09, 6.09, 3.83, 3.32, 6.62, 3.71)
boxplot(x, horizontal = TRUE, col="lightgrey")
hist(x, breaks = "freedman-diaconis", col = "lightgrey",
border = "black")
m=mean(x)
mo=mlv(x, method = "parzen", bw="SJ", kernel =
"epanechnikov")
s=sd(x)
skpI= (m-mo)/s
cat("Coeficiente de Pearson tipo 1 = ",skpI,"n")
library(boot)
b<-boot(data=x, function(x, i) {(mean(x[i])- mlv(x[i],
method = »parzen«, bw=bw.nrd0(x), kernel =
»epanechnikov«))/sd(x[i])}, R=1000)
b
plot(b)
boot.ci(b, conf=0.90, type=c(»bca«,»norm«), digits = 4)
```

Coeficiente de Pearson tipo 1 = 0.486

$$A_{P_1} = \frac{\bar{x} - mo}{s_{n-1}} = \frac{4.974 - 4.075}{1.851} = 0.486$$

El intervalo arrojado por R incluye el 0 por el método bca [-0.295, 1.027], pero no por el método normal [0.005, 1.708]. Se mantiene la hipótesis nula de simetría con un nivel de significación del 10%, ya que el método de percentiles corregidos de sesgo y estandarizados es más adecuado que el normal con muestras pequeñas (Mokhtar, Yusof & Sapiri, 2023).

Para tener referencias interpretativas en relación con la distribución normal que es simétrica y de colas medias, se hicieron estimaciones por intervalos con un nivel de confianza al 90 %, 95 % y 99 % por muestreo repetitivo con reemplazamiento (Tabla

2). Se generaron 32 muestras-población con distribución normal estándar de tamaño finito y rango de $z = -3.5$ ($F_z[z = -3.5] = \Phi(-3.5) = 0.0002$) a 3.5 ($F_z[z = 3.5] = \Phi(3.5) = 0.9998$). Se optó por muestras-población de n elementos normalmente distribuidos estrictamente simétricas para garantizar un valor poblacional nulo. Se adoptó el rango de -3.5 a 3.5 , ya que es muy raro que aparezcan datos a más de tres desviaciones estándar de la media aritmética en una distribución normal y corresponde a una amplitud total de ocho veces el rango intercuartílico y unas siete veces la desviación estándar, como considera la regla de la Universidad de Rice (Lane, 2015) para definir el número de intervalos de clase en el histograma ($k = [2 \times n^{1/3}]$) en relación con las reglas de Freedman y Diaconis (1981) y Scott (1979) y la distribución normal (Moral, 2023).

Los tamaños poblacionales N variaron de 10 a 200 con incrementos de 10 elementos, de 220 a 400 con incrementos de 20, de 450 a 500 con incremento de 50 y se termina con un tamaño de 1000. Se dividió el rango de probabilidades acumulativas por el tamaño muestral, $\Delta P = (0.999767371 - 0.000232629) / (N - 1)$. Los datos poblacionales se obtuvieron con la función probit: $z_k = \Phi^{-1}(P_k = 0.000232629 + (k - 1) \times \Delta P)$, donde $k = 1, 2, \dots, N$. Por ejemplo, para la población de tamaño 10: $\Delta P = 0.999534742 / 9 = 0.111059416$.

$Z_{(N=10)} = \{z_1 = -3.5, z_2 = -1.219685581, z_3 = -0.76427577, z_4 = -0.430514044, z_5 = -0.139644873, z_6 = 0.139644873, z_7 = 0.430514044, z_8 = 0.76427577, z_9 = 1.219685581, z_{10} = 3.5\}$.

De cada una de estas 32 muestras-población se extrajeron al azar 10 000 muestras aleatorias por muestreo repetitivo con reposición para calcular el error estándar y los intervalos de confianza al 90 %, 95 % y 99 % del coeficiente de asimetría de Pearson 1. Se usó la moda de Parzen (1962) con el kernel de Epanechnikov (1969) y ancho de banda de Silverman (1986). Las instrucciones para el programa R son las siguientes, concretadas para la muestra-población de 10 datos. El resultado con las 32 muestras-población se muestra en la Tabla 2.

```

library(modeest)
x<-c(-3.5, -1.219685581, -0.76427577, -0.430514044, -0.139644873, 0.139644873, 0.430514044,
0.76427577, 1.219685581, 3.5)
m=mean(x)
mo=mlv(x, method = "parzen", bw=bw.nrd0(x), kernel = "epanechnikov")
s=sd(x)
skpI=(m-mo)/s
cat("Coeficiente de Pearson tipo 1 = ",skpI,"n")
library(boot)
b<-boot(data=x, function(x, i) {(mean(x[i])- mlv(x[i], method = "parzen", bw=bw.nrd0(x), kernel =
"epanechnikov"))/sd(x[i])}, R=1000)
b
plot(b)
boot.ci(b, conf=c(0.90, 0.95, 0.99), type= "bca", digits = 4)

```

Tabla 2

Intervalos de confianza por muestreo repetitivo para A_{p1} con la moda de Parzen y el método de percentiles corregidos de sesgo y acelerados

N	A_{p1}	Sesgo	EE	IC al 90%		IC al 95%		IC al 99%	
				LI	LS	LI	LS	LI	LS
10	0.0061	-0.010	0.381	-0.600	0.609	-0.706	0.702	-0.942	0.876
20	0.0061	-0.012	0.382	-0.590	0.606	-0.705	0.710	-0.905	0.901
30	-0.0054	0.000	0.396	-0.652	0.647	-0.744	0.736	-0.948	0.890
40	-0.0319	0.026	0.393	-0.693	0.597	-0.786	0.698	-1.023	0.890
50	0.0059	0.001	0.390	-0.651	0.639	-0.748	0.726	-0.912	0.895
60	0.0060	-0.005	0.383	-0.636	0.641	-0.736	0.737	-0.935	0.910
70	-0.0085	0.004	0.381	-0.641	0.636	-0.741	0.738	-0.903	0.932
80	0.0328	-0.029	0.374	-0.577	0.689	-0.686	0.783	-0.833	0.936
90	-0.0063	0.009	0.370	-0.641	0.613	-0.733	0.712	-0.922	0.878
100	-0.0064	0.010	0.354	-0.625	0.575	-0.722	0.686	-0.900	0.860
110	0.0064	0.001	0.357	-0.607	0.605	-0.715	0.706	-0.883	0.860
120	0.0065	-0.011	0.351	-0.575	0.620	-0.684	0.717	-0.852	0.889
130	-0.0065	0.008	0.343	-0.610	0.559	-0.707	0.668	-0.871	0.841
140	-0.0065	0.007	0.338	-0.578	0.570	-0.674	0.676	-0.872	0.867
150	0.0001	0.006	0.336	-0.579	0.561	-0.679	0.668	-0.840	0.844
160	-0.0066	0.002	0.332	-0.577	0.551	-0.687	0.655	-0.838	0.823
170	0.0066	-0.008	0.325	-0.516	0.585	-0.638	0.682	-0.811	0.838
180	0.0041	-0.005	0.319	-0.516	0.565	-0.627	0.671	-0.812	0.839
190	0.0041	-0.008	0.319	-0.505	0.575	-0.610	0.681	-0.774	0.871
200	0.0066	-0.010	0.314	-0.505	0.554	-0.618	0.658	-0.784	0.825
220	-0.0066	0.006	0.304	-0.529	0.504	-0.627	0.599	-0.792	0.772

N	A_{p1}	Sesgo	EE	IC al 90%		IC al 95%		IC al 99%	
				LI	LS	LI	LS	LI	LS
240	-0.0067	0.007	0.303	-0.532	0.486	-0.654	0.594	-0.797	0.770
260	-0.0002	-0.003	0.295	-0.478	0.513	-0.572	0.621	-0.757	0.792
280	0.0001	0.000	0.286	-0.474	0.474	-0.580	0.585	-0.752	0.761
300	0.0042	-0.003	0.284	-0.460	0.499	-0.561	0.598	-0.728	0.755
320	-0.0001	-0.001	0.279	-0.458	0.462	-0.544	0.568	-0.729	0.767
340	0.0001	-0.001	0.276	-0.461	0.460	-0.559	0.562	-0.730	0.734
360	0.0001	0.004	0.273	-0.463	0.446	-0.557	0.543	-0.732	0.742
380	0.0042	-0.004	0.273	-0.441	0.462	-0.540	0.559	-0.704	0.738
400	-0.0002	-0.002	0.266	-0.434	0.444	-0.525	0.529	-0.712	0.706
450	0.0001	-0.003	0.257	-0.423	0.427	-0.506	0.504	-0.677	0.676
500	0.0001	0.001	0.254	-0.424	0.414	-0.507	0.498	-0.676	0.673
1000	0.0001	0.002	0.214	-0.354	0.350	-0.422	0.418	-0.554	0.550

Nota. N = tamaño de la población, A_{p1} = coeficiente de asimetría de Pearson 1, EE = error estándar, IC = intervalo de confianza, LI = límite inferior y LS = límite superior del intervalo de confianza. Kernel de Epanechnikov y ancho de banda de Silverman (1986) para el cálculo de la moda de Parzen (1962). Número de extracciones: 10 000.

La moda de Grenander

La moda también se puede obtener por el estimador de Grenander (1965), resultando unos intervalos de confianza para el coeficiente de asimetría mucho más eficientes, esto es, con menor amplitud. Se ordenan los n datos muestrales en sentido ascendente.

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$\widehat{Mo} = M_{p,k}(n) = \frac{\frac{1}{2} \sum_{i=1}^{n-k} \frac{x_{(i+k)} + x_{(i)}}{[x_{(i+k)} - x_{(i)}]^p}}{\sum_{i=1}^{n-k} \frac{1}{[x_{(i+k)} - x_{(i)}]^p}}$$

$$1 < p < k; 2 \leq k < n; p, k \in \mathbb{R}^+$$

Si $k > 2p$, $M_{p,k}(n)$ converge a una distribución normal. Si la distribución es simétrica, como la distribución normal, el estimador $M_{p,k}(n)$ es insesgado, pero si la distribución es asimétrica, el estimador $M_{p,k}(n)$ presenta sesgo. La forma de reducir el sesgo es incrementar el valor de p . Así, una opción es adoptar un valor central para k :

$$k = \lfloor n/2 \rfloor + 1$$

$$p = \lfloor n/4 \rfloor; k > 2p$$

Retomando el ejemplo previo con la muestra aleatoria de 16 datos, las instrucciones para calcular A_{p1} con la moda de Grenander (1965) son las siguientes:

```
library(modeest)
x<-c(5.26, 7.58, 1.43, 2.90, 6.37, 8.51, 6.31, 4.63,
4.92, 4.02, 4.09, 6.09, 3.83, 3.32, 6.62, 3.71)
m=mean(x)
mo=mlv(x, method="grenander", bw=NULL,
k=floor(length(x)/2)+1, p=floor(length(x)/4))
cat("Moda = ",mo,"\n")
s=sd(x)
skpI=(m-mo)/s
cat("Coeficiente de Pearson tipo 1 = ",skpI,"\n")
library(boot)
b<-boot(data=x, function(x, i) {(mean(x[i])-
mlv(x[i], method="grenander", bw=NULL, k=
floor(length(x[i])/2)+1, p=floor(length(x[i])/4)))/
sd(x[i])}, R=1000)
b
plot(b)
boot.ci(b, conf=0.90, type=c("norm","bca"),
digits=3)
```

El resultado arrojado muestra una clara simetría tanto por el método bca como por el normal: mo

= 4.973, $A_{p1} = (4.974 - 4.973) / 1.851 = 0.001$, IC bootstrap normal al 90 % (-0.378, 0.263) e IC bootstrap bca al 90 % (-0.408, 0.255).

Se calculan intervalos de confianza para una asimetría nula mediante simulación paramétrica (método normal) con la extracción con reemplazamiento de 10 000 muestras aleatorias a partir de cada de las 33 muestras-población de datos normalmente distribuidos (Tabla 3). El cálculo se hace por medio del programa R, las instrucciones son las siguientes:

```
library(modeest)
x<-c(-3.5, -1.219685581, -0.76427577, -0.430514044, -0.139644873,
0.139644873, 0.430514044, 0.76427577, 1.219685581, 3.5)
m=mean(x10)
mo= mlv(x10, method = "grenander", bw = NULL, k =
floor(length(x10)/2) + 1, p = floor(length(x10)/4))
s=sd(x10)
skpI= (m-mo)/s
cat("Coeficiente de Pearson tipo 1 = ",skpI,"n")
library(boot)
b<-boot(data=x10, function(x10, i) {(mean(x10[i])- mlv(x10[i],
method = »grenander«, bw = NULL, k = floor(length(x10[i])/2) + 1, p
= floor(length(x10[i])/4)))/sd(x10[i])}, R=10000)
b
plot(b)
boot.ci(b, conf=c(0.90, 0.95, 0.99), type= »norm«, digits = 3)
```

Tabla 3

Intervalos de confianza por muestreo repetitivo y método normal para A_{p1} con moda de Grenander

N	A_{p1}	Sesgo	EE	IC al 90%		IC al 95%		IC al 99%	
				LI	LS	LI	LS	LI	LS
10	0	0.002	0.174	-0.288	0.285	-0.343	0.340	-0.451	0.447
20	0	0.002	0.204	-0.337	0.333	-0.402	0.397	-0.527	0.523
30	0	0.001	0.189	-0.311	0.310	-0.370	0.369	-0.486	0.485
40	0	-0.001	0.189	-0.310	0.313	-0.370	0.372	-0.487	0.489
50	0	-0.002	0.181	-0.296	0.299	-0.353	0.356	-0.465	0.468
60	0	-0.002	0.180	-0.294	0.298	-0.351	0.355	-0.462	0.466
70	0	0.003	0.177	-0.294	0.287	-0.349	0.343	-0.458	0.452
80	0	-0.0001	0.176	-0.290	0.290	-0.345	0.345	-0.453	0.454
90	0	-0.0005	0.173	-0.284	0.285	-0.338	0.339	-0.445	0.446
100	0	-0.0002	0.170	-0.279	0.280	-0.333	0.333	-0.437	0.438
110	0	-0.001	0.167	-0.274	0.275	-0.326	0.328	-0.429	0.431
120	0	0.0004	0.165	-0.272	0.272	-0.324	0.324	-0.426	0.425
130	0	0.002	0.163	-0.269	0.266	-0.321	0.317	-0.421	0.417
140	0	-0.002	0.163	-0.266	0.269	-0.317	0.320	-0.417	0.420
150	0	-0.001	0.160	-0.263	0.265	-0.314	0.315	-0.413	0.414
160	0	<0.001	0.159	-0.261	0.261	-0.311	0.311	-0.409	0.409
170	0	0.001	0.155	-0.256	0.254	-0.305	0.303	-0.400	0.398
180	0	0.002	0.153	-0.254	0.250	-0.302	0.298	-0.396	0.392
190	0	0.001	0.153	-0.253	0.251	-0.301	0.299	-0.395	0.393
200	0	-0.001	0.151	-0.248	0.250	-0.295	0.297	-0.389	0.391

N	A_{P1}	Sesgo	EE	IC al 90%		IC al 95%		IC al 99%	
				LI	LS	LI	LS	LI	LS
220	0	0.003	0.148	-0.245	0.240	-0.292	0.287	-0.383	0.378
240	0	0.001	0.145	-0.239	0.237	-0.285	0.282	-0.374	0.371
260	0	<0.001	0.144	-0.237	0.237	-0.282	0.282	-0.371	0.371
280	0	0.002	0.143	-0.237	0.233	-0.281	0.278	-0.369	0.366
300	0	-0.002	0.138	-0.226	0.229	-0.269	0.273	-0.354	0.358
320	0	-0.001	0.137	-0.224	0.227	-0.267	0.270	-0.351	0.354
340	0	-0.001	0.134	-0.219	0.221	-0.261	0.264	-0.344	0.346
360	0	-0.0002	0.132	-0.217	0.218	-0.259	0.259	-0.340	0.341
380	0	0.002	0.133	-0.220	0.217	-0.262	0.259	-0.344	0.341
400	0	-0.001	0.132	-0.216	0.218	-0.257	0.259	-0.339	0.341
450	0	<0.001	0.126	-0.207	0.207	-0.247	0.246	-0.324	0.324
500	0	0.001	0.123	-0.204	0.202	-0.243	0.241	-0.319	0.317
1000	0	-0.001	0.102	-0.167	0.170	-0.199	0.202	-0.262	0.265

Nota. N = tamaño de la población, A_{P1} = coeficiente de asimetría de Pearson 1, EE = error estándar, IC = intervalo de confianza, LI = límite inferior y LS = límite superior del intervalo de confianza.

Medida robusta de asimetría basada en la moda de Bickel

La asimetría de Bickel (2002) a nivel poblacional se define como el complemento del valor en la función de distribución acumulativa de la moda poblacional y se puede denotar por A_B . Su valor varía de -1 a 1 y 0 indica simetría.

$$A_B = 1 - F_X(Mo)$$

A nivel muestral requiere estimar la moda. Dada una muestra de tamaño n , A_B se calcula mediante la siguiente fórmula, donde $\#$ es la cardinalidad o el número de veces que se cumple la condición entre los n elementos de la muestra.

$$\hat{A}_B = 1 - 2 \times \frac{\#(x = \widehat{Mo}) + \#(x < \widehat{Mo})}{n}$$

La moda se estima por medio del método de semimuestras reiterativas de Bickel (2002). Se trata de un procedimiento no paramétrico y robusto (Ruzankin, 2022). Este proceso reiterativo consiste en dividir la muestra total en submuestras con $[n/2]+1$ elementos y se busca la que tenga el rango mínimo. Se repite el procedimiento con esta submuestra y las siguientes hasta obtener una submuestra de dos elementos y con su pro-

medio se obtiene la moda o hasta quedarse con un solo elemento que es la moda. Véase el cálculo en la muestra de 16 elementos previamente utilizada.

Paso 1: Se ordenan los datos de la muestra total o inicial en sentido ascendente:

$s_1 = \{1.43, 2.9, 3.32, 3.71, 3.83, 4.02, 4.09, 4.63, 4.92, 5.26, 6.09, 6.31, 6.37, 6.62, 7.58, 8.51\}$.

Cardinalidad de la muestra s_1 : $n = 16$.

$k = [n/2] + 1 = [16/2] + 1 = 9 \rightarrow$ Se divide en ocho submuestras de nueve elementos ordenados, se calcula el rango de cada submuestra y se busca el rango mínimo.

$Min(4.92 - 1.43 = 3.49, 5.26 - 2.9 = 2.36, 6.09 - 3.32 = 2.77, 6.31 - 3.71 = 2.6, 6.37 - 3.83 = 2.54, 6.62 - 4.02 = 2.6, 7.58 - 4.09 = 3.49, 8.51 - 4.63 = 3.88) = 2.36$.

Paso 2: Se retiene la submuestra con el rango mínimo: $s_2 = \{2.9, 3.32, 3.71, 3.83, 4.02, 4.09, 4.63, 4.92, 5.26\}$.

Cardinalidad de la muestra s_2 : $n = 9$.

$k = [n/2] + 1 = [9/2] + 1 = 5 \rightarrow$ Se divide en cinco submuestras de cinco elementos ordenados, se calcula el rango de cada submuestra y se busca el rango mínimo.

$Min(4.02 - 2.9 = 1.12, 4.09 - 3.32 = 0.77, 4.63 - 3.71 = 0.92, 4.92 - 3.83 = 1.09, 5.26 - 4.02 = 1.24) = 0.77$.

Paso 3: Se retiene la submuestra con el rango mínimo: $s_3 = \{3.32, 3.71, 3.83, 4.02, 4.09\}$.

Cardinalidad de la muestra s_3 : 5.

$k = \lfloor n/2 \rfloor + 1 = \lfloor 5/2 \rfloor + 1 = 3 \rightarrow$ Se divide en tres submuestras de tres elementos, se calcula el rango de cada submuestra y se busca el rango mínimo.

$Min(3.83 - 3.32 = 0.51, 4.02 - 3.71 = 0.31, 4.09 - 3.83 = 0.26) = 0.26$.

Paso 4: Se retiene la submuestra con el rango mínimo: $s_4 = \{3.83, 4.02, 4.09\}$.

Cardinalidad de la muestra s_4 : 3.

$k = \lfloor n/2 \rfloor + 1 = \lfloor 3/2 \rfloor + 1 = 2 \rightarrow$ Se divide en dos submuestras de dos elementos, se calcula el rango de cada submuestra y se busca el rango mínimo.

$Min(4.02 - 3.83 = 0.19, 4.09 - 4.02 = 0.07) = 0.07$.

Paso 5: Se retiene la submuestra con el rango mínimo: $s_5 = \{4.02, 4.09\}$.

Cardinalidad de la muestra s_5 : 2.

$k = \lfloor n/2 \rfloor + 1 = \lfloor 2/2 \rfloor + 1 = 2 \rightarrow$ una muestra de dos elementos.

La moda es el promedio de los dos elementos: $Mo = (4.02 + 4.09)/2 = 4.055$.

Una vez estimada la moda en esta muestra aleatoria de 16 datos por el método de semimuestras reiterativas, se puede calcular la asimetría robusta de Bickel (2002).

$$\hat{A}_B = 1 - 2 \times \frac{\frac{\#(x = \widehat{Mo})}{2} + \#(x < \widehat{Mo})}{n}$$

$$= 1 - 2 \times \frac{0}{2} + 6 = 1 - 0.75 = 0.25$$

Cuando se estima el coeficiente por intervalo con un nivel de confianza del 90 % mediante el procedimiento de percentiles corregidos de sesgo y acelerados, se observa que el 0 está incluido, por lo que se concluye que la muestra es simétrica con un nivel de significación a dos colas del 10 %.

$P(LI_{A_B} \leq A_B \leq LS_{A_B}) = P(-0.5045 \leq A_B \leq 0.6875) = 0.90$

$$0 \in [-0.5045, 0.6875] \rightarrow A_B = 0$$

La moda por semimuestras reiterativas, el coeficiente de asimetría y el intervalo de confianza se pueden calcular con el programa R mediante las siguientes instrucciones. Como resultado es el siguiente: $n = 16$,

moda = 4.055, $A_B = 0.25$, sesgo = -0.0618, error estándar = 0.3812, IC al 90% (-0.5045, 0.6875) por el método bca.

```
library(modeest)
x<-c(5.26, 7.58, 1.43, 2.90, 6.37, 8.51, 6.31, 4.63,
4.92, 4.02, 4.09, 6.09, 3.83, 3.32, 6.62, 3.71)
ss <- length(x)
cat(«Tamaño muestral = »,ss,«\n»)
Mo=mlv(x, method = "hsm", bw = NULL, k =
floor(ss/2) + 1, tie.action = "mean", tie.limit =
0.05)
cat("Moda = ",Mo,"n")
calcular_A <- function(x, Mo) {num_valores_Mo
<- sum(x == Mo)
num_valores_menores_Mo <- sum(x < Mo)
A <- 1 - 2 * ((num_valores_Mo / 2 + num_valores_
menores_Mo) / ss)
return(A)}
AB <- calcular_A(x, Mo)
cat("Coeficiente robusto de asimetría de Bickel =
",AB,"n")
library(boot)
b <- boot(data = x, statistic = function(x, i) {
num_valores_Mo <- sum(x[i] == mlv(x[i],
method = "hsm", bw = NULL, k =
floor(length(x[i])/2) + 1, tie.action = "mean", tie.
limit = 0.05))
num_valores_menores_Mo <- sum(x[i] <
mlv(x[i], method = "hsm", bw = NULL, k =
floor(length(x[i])/2) + 1, tie.action = "mean", tie.
limit = 0.05))
A <- 1 - 2 * ((num_valores_Mo / 2 + num_valores_
menores_Mo) / length(x[i]))
return(A)}, R = 1000)
b
plot(b)
boot.ci(b, conf = 0.90, type = «bca», digits = 4)
```

En la Tabla 4, se presentan intervalos al 90 %, 95 % y 99 % para muestras de diversos tamaños de datos normalmente distribuidos correspondientes a una asimetría nula, obtenidos por muestreo repetitivo con reemplazamiento mediante el método no paramétrico de percentiles corregidos de sesgo y acelerados (bca), salvo el último con la muestra de 1000 datos que rebasa la capacidad de cálculo del programa, por lo que se acudió al método paramétrico o normal. Se opta por el método no paramétrico al ser más eficiente en general con

este estadístico, especialmente con muestras pequeñas. Para los cálculos, se usan las 33 muestras-población generadas y se modifica la última línea del guion de instrucciones: `boot.ci(b, conf = c(0.90, 0.95, 0.99), type = »bca«, digits = 3)`. Las gráficas (histograma y gráfica

cuantil-cuantil) muestran que las distribuciones extraídas de muestras-población pequeñas claramente se alejan de la normalidad, pero se observa una convergencia a la distribución normal según se incrementa el tamaño muestral.

Tabla 4

Intervalos de confianza para muestras normalmente distribuidas de distintos tamaños por el método de percentiles corregidos de sesgo y acelerados

N	A _B	Sesgo	EE	Método	IC al 90 %		IC al 95 %		IC al 99 %	
					LI	LS	LI	LS	LI	LS
10	0	0.187	0.362	bca	-0.800	0.300	-0.800	0.400	-0.800	0.600
20	0	0.050	0.333	bca	-0.650	0.450	-0.750	0.550	-0.850	0.700
30	0	0.062	0.308	bca	-0.633	0.400	-0.700	0.500	-0.810	0.667
40	0	0.033	0.293	bca	-0.550	0.425)	-0.625	0.525	-0.750	0.700
50	0	0.032	0.283	bca	-0.540	0.420	-0.620	0.520	-0.720	0.660
60	0	0.046	0.276	bca	-0.533	0.400	-0.600	0.483	-0.733	0.633
70	0	0.033	0.267	bca	-0.500	0.400	-0.571	0.471	-0.700	0.629
80	0	0.033	0.260	bca	-0.488	0.388	-0.563	0.463	-0.692	0.613
90	0	0.018	0.253	bca	-0.444	0.400	-0.522	0.478	-0.644	0.611
100	0	0.017	0.249	bca	-0.440	0.390	-0.510	0.470	-0.650	0.620
110	0	0.013	0.246	bca	-0.436	0.391	-0.509	0.464	-0.627	0.600
120	0	0.025	0.244	bca	-0.442	0.383	-0.508	0.458	-0.615	0.592
130	0	0.020	0.242	bca	-0.423	0.392	-0.500	0.469	-0.608	0.585
140	0	0.018	0.236	bca	-0.429	0.369	-0.493	0.443	-0.620	0.571
150	0	0.018	0.234	bca	-0.413	0.375	-0.473	0.447	-0.582	0.567
160	0	0.016	0.232	bca	-0.419	0.363	-0.481	0.431	-0.600	0.565
170	0	0.018	0.227	bca	-0.400	0.365	-0.459	0.435	-0.571	0.557
180	0	0.019	0.228	bca	-0.414	0.350	-0.482	0.417	-0.600	0.544
190	0	0.014	0.224	bca	-0.395	0.358	-0.468	0.427	-0.576	0.549
200	0	0.008	0.221	bca	-0.380	0.360	-0.445	0.430	-0.555	0.555
220	0	0.008	0.221	bca	-0.373	0.364	-0.432	0.432	-0.550	0.549
240	0	0.008	0.216	bca	-0.375	0.354	-0.438	0.413	-0.553	0.521
260	0	0.012	0.215	bca	-0.371	0.342	-0.438	0.413	-0.563	0.538
280	0	0.010	0.212	bca	-0.373	0.339	-0.435	0.403	-0.537	0.518
300	0	0.010	0.208	bca	-0.363	0.333	-0.420	0.397	-0.513	0.513
320	0	0.007	0.206	bca	-0.359	0.334	-0.416	0.394	-0.528	0.512

N	A_B	Sesgo	EE	Método	IC al 90 %		IC al 95 %		IC al 99 %	
					LI	LS	LI	LS	LI	LS
340	0	0.014	0.204	bca	-0.359	0.321	-0.418	0.386	-0.538	0.491
360	0	0.009	0.201	bca	-0.350	0.322	-0.411	0.381	-0.504	0.487
380	0	0.008	0.202	bca	-0.347	0.324	-0.400	0.384	-0.490	0.495
400	0	0.008	0.200	bca	-0.335	0.333	-0.398	0.392	-0.504	0.488
450	0	0.009	0.195	bca	-0.333	0.318	-0.393	0.380	-0.487	0.470
500	0	0.004	0.192	bca	-0.322	0.322	-0.380	0.376	-0.482	0.476
1000	0	0.002	0.172	norm	-0.286	0.282	-0.340	0.336	-0.446	0.442

Nota. N = tamaño de la población, A_B = coeficiente de asimetría de Bickel, EE = error estándar, método: bca = percentiles corregidos de sesgo y acelerados, norm = normal o gaussiano, IC = intervalo de confianza, LI = límite inferior y LS = límite superior del intervalo de confianza.

Si la variable es ordinal y su distribución unimodal, la mejor opción es la asimetría robusta de Bickel (2002). El método para estimar la moda es *mfv*, que devuelve el o los valores más frecuentes de un vector numérico dado. Véase un ejemplo. Sea X la preferencia sexual con cinco categorías: 1 = heterosexual, 2 = preferentemente he-

terosexual, 3 = bisexual, 4 = preferentemente homosexual y 5 = homosexual. Se extrajo de una población de hombres jóvenes una muestra de 20 participantes. ¿La muestra es simétrica?

Se calcula la moda, A_B y el intervalo de confianza por medio de las siguientes instrucciones:

```
library(modeest)
x<-c(5, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 5, 4, 1)
ss <- length(x)
cat("Tamaño muestral = ",ss,"\n")
Mo=mlv(x, method = "mfv")
cat("Moda = ",Mo,"\n")
calcular_A <- function(x, Mo) {num_valores_Mo <- sum(x == Mo)
num_valores_menores_Mo <- sum(x < Mo)
A <- 1 - 2 * ((num_valores_Mo / 2 + num_valores_menores_Mo) / ss)
return(A)}
AB <- calcular_A(x, Mo)
cat("Coeficiente robusto de asimetría de Bickel = ",AB,"\n")
library(boot)
b <- boot(data = x, statistic = function(x, i) {
num_valores_Mo <- sum(x[i] == mlv(x[i], method = "mfv"))
num_valores_menores_Mo <- sum(x[i] < mlv(x[i], method = "mfv"))
A <- 1 - 2 * ((num_valores_Mo / 2 + num_valores_menores_Mo) / length(x[i]))
return(A)}, R = 1000)
b
plot(b)
boot.ci(b, conf = 0.90, type = «bca», digits = 4)
```

El resultado evidencia que distribución poblacional de la cual se extrajo la muestra aleatoria es asimétrica positiva. El índice salió positivo y mayor que 0

con un nivel de significación del 10 %: $A_B = 0.2$, sesgo = -0.002, error estándar = 0.090, IC al 90 % (0.05, 0.35).

Discusión

La medición de la asimetría basada en la moda tiene un requisito fundamental que es la unimodalidad. La función de densidad de probabilidad (variable continua) o la función de masa de probabilidad (variable discreta) de la variable aleatoria tiene que presentar un pico. Estas medidas no son válidas para distribuciones bimodales o multimodales, salvo que sea dos valores consecutivos de frecuencia máxima. En ese caso, la moda es su media. La medición de Pearson (1894) está desarrollada para variable cuantitativa. Requiere el cálculo de la media aritmética y la desviación estándar, por lo que no es adecuada para variables ordinales. En cambio, la medida de Bickel (2002) es aplicable tanto a variables cuantitativas como ordinales. Ninguna de estas dos medidas se aplica a variables cualitativas, pero sí existe una propuesta de medición de asimetría con estas variables (Moral, 2022).

De forma genérica, se sugiere que valores de 0 más menos una décima reflejan simetría con tamaños de muestra medios. Valores por debajo de -0.10 pueden indicar asimetría negativa y valores mayores que 0.10 pueden reflejar asimetría positiva. Con tamaños de muestra pequeños, el error de una décima será mayor y, con tamaños de muestra grandes, será menor debido a que la precisión de la estimación se incrementa con el aumento del tamaño de la muestra (Pakgohar & Mehrannia, 2023). Esta regla es del todo inválida para las medidas de asimetría basadas en la moda, como evidencian las tablas presentadas en este trabajo. El intervalo debe ser mayor.

El uso de la asimetría de Pearson (1894) con la moda de Grenander (1965) proporciona los intervalos con menor amplitud. Si la variable es cuantitativa, su distribución no presenta casos atípicos o colas pesadas y la distribución es unimodal o bimodal con dos valores modales consecutivos, es la mejor opción. Los intervalos basados en muestreo repetitivo con reemplazamiento se pueden obtener por el método normal, especialmente si el parámetro k es más del doble que el parámetro p y su valor es grande. La regla de un medio del tamaño de la muestra redondeado a la baja más uno para k y un cuarto del tamaño de la muestra redondeado a la baja para p garantizan esta condición. Si la

muestra es pequeña, el método de percentiles corregidos de sesgo y acelerados proporciona un intervalo de confianza con una amplitud más pequeña que el normal (Mokhtar et al., 2023).

La moda de Parzen (1962) es peor opción que la de Grenander (1965). Con esta moda basada en el máximo de la densidad estimada, el intervalo de confianza por el método de percentiles corregidos de sesgo y acelerados da mejor resultado que el obtenido por el método normal, sobre todo con muestras pequeñas. La distribución bootstrap se aleja de la normalidad con muestras pequeñas, pero converge a la distribución normal con muestras grandes. Si el tamaño de muestra es muy grande ($n \geq 1000$), entonces, el método normal resulta la mejor opción.

La asimetría robusta de Bickel (2002) parece mejor opción que la asimetría de Pearson con moda de Parzen cuando la variable es cuantitativa. Con variables ordinales es la mejor opción de las tres medidas, ya que solo requiere que los datos sean ordinales y tengan una moda o dos modas consecutivas. Para estimar la moda no se usa el método de Bickel (2002), sino el método del valor muestral de frecuencia máxima. Como en el caso previo, la distribución bootstrap se aleja de la normalidad con muestras pequeñas, pero converge a la distribución normal con muestras grandes. La estimación por percentiles corregidos de sesgo y acelerados es buena opción con esta medida, especialmente si la variable es ordinal o discreta con un rango muy limitado (Mokhtar et al., 2023).

Se concluye que la medida de asimetría de Pearson con la moda estimada por el método de Grenander (1965) y el intervalo de confianza de muestreo repetitivo con reemplazamiento obtenido por el método normal es la mejor opción con variables cuantitativas y la asimetría de Bickel (2002) con la moda calculada por el método del valor con frecuencia máxima y el intervalo de confianza obtenido por el método de percentiles corregidos de sesgo y acelerados lo es para variables ordinales. Cabe señalar que la intención del presente estudio metodológico de simulación, centrado en el análisis estadístico de datos, es facilitar el uso de estas medidas que resultan sencillas de cálculo e interpretación con el programa R (Giorgi, Ceraolo & Mercatelli, 2022).

Referencias

- Bickel, D. R. (2002). Robust estimators of the mode and skewness of continuous data. *Computational Statistics & Data Analysis*, 39(2), 153-163. [https://doi.org/10.1016/S0167-9473\(01\)00057-3](https://doi.org/10.1016/S0167-9473(01)00057-3)
- Bono, R., Arnau, J., Alarcón, R. & Blanca, M. J. (2020). Bias, precision, and accuracy of skewness and kurtosis estimators for frequently used continuous distributions. *Symmetry*, 12(1), article 19, 1-17. <https://doi.org/10.3390/sym12010019>
- Canty, A. & Ripley, B. (2022). Boot: bootstrap R (S-Plus) functions. R package version 1.3-28. <https://cran.r-project.org/web/packages/boot/boot.pdf>
- Eberl, A. & Klar, B. (2020). Asymptotic distributions and performance of empirical skewness measures. *Computational Statistics & Data Analysis*, 146, article 106939. <https://doi.org/10.1016/j.csda.2020.106939>
- Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications*, 14(1), 153-158. <https://doi.org/10.1137/1114019>
- Freedman, D. & Diaconis, P. (1981). On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4), 453-476. <https://doi.org/10.1007/BF01025868>
- Giorgi, F. M., Ceraolo, C. & Mercatelli, D. (2022). The R language: an engine for bio-informatics and data science. *Life*, 12(5), article 648. <https://doi.org/10.3390/life12050648>
- Grenander, U. (1965). Some direct estimates of the mode. *Annals of Mathematical Statistics*, 36(1), 131-138. <https://doi.org/10.1214/aoms/1177700277>
- Guidoum, A. C. (2020). Kernel estimator and bandwidth selection for density and its derivatives: the kedd package. *arxiv*, article 2012.06102v1. <https://doi.org/10.48550/arXiv.2012.06102>
- Gupta, S. C. & Kapoor, V. K. (2020). Descriptive measures. In *Fundamentals of mathematical statistics, twelfth edition* (section 2, pp. 1-78). New Delhi: Sultan Chand & Sons.
- Henderson, D. J., Papadopoulos, A. & Parmeter, C. F. (2023). Bandwidth selection for kernel density estimation of fat-tailed and skewed distributions. *Journal of Statistical Computation and Simulation*, article 2173194, 1-26. <https://doi.org/10.1080/00949655.2023.2173194>
- Khan, A. A., Cheema, S. A., Hussain, Z. & Abdel-Salam, G. A. (2021). Measuring skewness: We do not assume much. *Scientia Iranica*, 28(6), 3525-3537. <https://doi.org/10.24200/SCI.2020.52306.2649>
- Lane, D. M. (2015). Histograms. En *Online statistics education: a multimedia course of study*. [https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_\(Lane\)/02%3A_Graphing_Distributions/2.04%3A_Histograms](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Lane)/02%3A_Graphing_Distributions/2.04%3A_Histograms)
- Mokhtar, S. F., Yusof, Z. M. & Sapiri, H. (2023). Confidence intervals by bootstrapping approach: a significance review. *Malaysian Journal of Fundamental and Applied Sciences*, 19(1), 30-42. <https://doi.org/10.11113/mjfas.v19n1.2660>
- Moral, J. (2022). Una medida de asimetría unidimensional para variables cualitativas. *Revista de Psicología (PUCP)*, 40(1), 519-551. <https://dx.doi.org/10.18800/psico.202201.017>
- Moral, J. (2023). Standardized distance from the mean to the median as a measure of skewness. *Open Journal of Statistics*, 13, 359-378. <https://dx.doi.org/10.4236/ojs.2023.133018>
- Pakgothar, A. & Mehrannia, H. (2023). Statistical rules in scientific reports (the basics). *Iranian Journal of Diabetes and Obesity*, 15, article 12205. <https://doi.org/10.18502/ijdo.v15i1.12205>
- Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3), 1065-1076. <https://doi.org/10.1214/aoms/1177704472>
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. I. On the dissection of asymmetrical frequency curves. *Philosophical Transactions of the Royal Society of London A*, 185, 71-110. <https://doi.org/10.1098/rsta.1894.0003>
- Poncet, P. (2022). Package 'modeest'. *Mode estimation*. <https://cran.r-project.org/web/packages/modeest/modeest.pdf>
- Ruzankin, P. S. (2022). A class of nonparametric mode estimators. *Communications in Statistics - Simulation and Computation*, 51(6), 3291-3304. <https://doi.org/10.1080/03610918.2019.1711410>

- Sarka, D. (2021). Descriptive statistics. In: *Advanced Analytics with Transact-SQL* (pp. 3-29). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-7173-5_1
- Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3), 605-610. <https://doi.org/10.1093/biomet/66.3.605>
- Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3), 683-690. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>
- Shi, J., Luo, D., Wan, X., Liu, Y., Liu, J., Bian, Z. & Tong, T. (2020). Detecting the skewness of data from the sample size and the five-number summary. *arXiv*, article 2010.05749. <https://doi.org/10.48550/arXiv.2010.05749>
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.