La simetría no siempre es una campana de Gauss

Mauricio Fuentes Alburquenque
Programa de Bioestadística, Escuela de Salud Pública, Facultad de Medicina
Universidad de Chile

Por lo general en clases de estadística descriptiva y en los libros que abordan este tópico se dice que si en un conjunto de datos la media, la mediana y la moda coinciden, entonces se trata de una distribución simétrica y esta simetría se asocia a la conocida "campana de Gauss". Más aún, se induce a pensar que mientras más simétrica es una distribución más se aproxima a la mencionada forma de campana, llegando la curva gaussiana a representar la simetría máxima o ideal.

Si bien es cierto que cuando estas tres medidas de resumen coinciden la distribución es simétrica, lo inverso no siempre es cierto, a pesar de lo cual muchos, por costumbre, caemos en el error de considerarlo como tal. Un ejemplo de esto ocurrió durante este año en la asignatura de Bioestadística para la carrera de Medicina, donde el equipo docente –del cual formo parte– incluyó la siguiente pregunta en una de las evaluaciones:

Seleccione la afirmación que es FALSA:

a) *La mediana es una mejor medida de tendencia central para datos con distribuciones muy asimétricas.*
b) *Un conjunto de datos siempre presenta una moda.*
c) *Si los datos tienen una distribución simétrica, entonces la media, mediana y moda coinciden.*
d) *La media es sensible a datos en los extremos de la distribución.*

La pauta de evaluación, que se hizo llegar a las y los estudiantes, indicaba que la respuesta es la letra b (supuestamente la única afirmación falsa). Sin embargo, un tiempo después una estudiante de la carrera nos hizo llegar la siguiente consulta:

Hola profe!
Soy ... Escribo para apelar al control, en la pregunta "Seleccione la afirmación que es FALSA", marqué la respuesta "Si los datos tienen una distribución simétrica, entonces la media, mediana y moda coinciden.", esta afirmación es falsa, contraejemplo: una muestra con dos datos, con valores 1 y -1, media=mediana=0, pero moda= 1,-1.
No marqué la alternativa que se indica por pauta, porque también se podría interpretar que quiere decir que siempre hay al menos una moda (lo cual es cierto).

Esto nos llevó a conversar el asunto al interior del equipo y no pudimos sino estar de acuerdo con la estudiante, a quien llamaremos **V**[1], respecto a que la alternativa c era también falsa[2]. No obstante, y tal como le respondimos, su ejemplo no fue el más acertado ya que corresponde a una distribución amodal, aunque no cabía duda que se entendía hacia dónde apuntaba.

No cuesta mucho imaginar distribuciones que sigan patrones parecidos al señalado por **V**, por ejemplo, como las mostradas en las figuras 1.a) y 1.b)[3]. En ambos casos la media y la mediana coinciden y se ubican en el valor central de la variable aleatoria X, pero las modas se ubican en valores equidistantes de las otras dos medidas. El coeficiente de asimetría en ambas distribuciones es cero[4], indicando que son simétricas aunque no gaussianas.

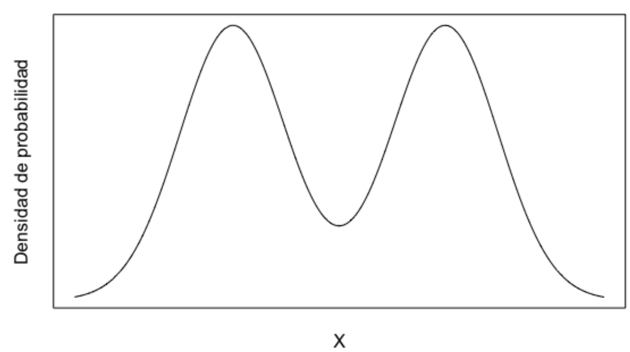

Figura 1.a): Distribución simétrica bimodal con las modas en valores intermedios de la variable aleatoria.

[1] Consulté a **V** y ella prefirió permanecer en el anonimato.
[2] Su última afirmación *"siempre hay al menos una moda (lo cual es cierto)"* es incorrecta, pero no es el foco de este artículo.
[3] Una cosa es imaginarlas y la otra es construir algo parecido a lo que se está imaginando. Las distribuciones mostradas se construyeron sumando ponderadamente dos funciones de densidad gaussianas $N(\mu = 0{,}3; \sigma = 0{,}1)$ y $N(\mu = 0{,}7; \sigma = 0{,}1)$ (Figura 1.a), y dos funciones de densidad $Beta(a = 1/2; b = 1)$ y $Beta(a = 1, b = 1/2)$ (Figura 1.b).
[4] Aunque hay varias versiones del coeficiente de asimetría, todas tienen como elemento principal el tercer momento central, es decir, la suma $\sum(x_i - \mu)^3$. Dado que la distribución es simétrica en torno a la media, todas las diferencias al cubo conservan el signo y por lo tanto se anulan, con lo que la suma es cero.

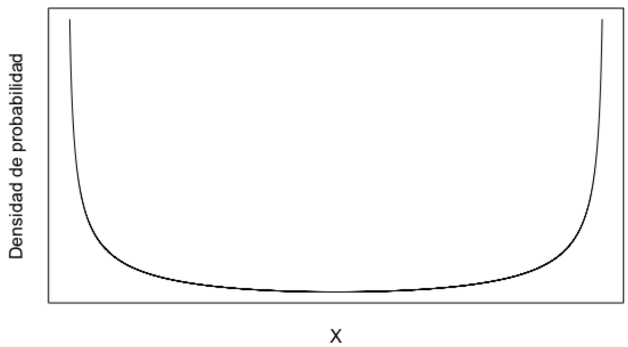

Figura 1.b): Distribución simétrica bimodal con las modas en los valores extremos de la variable aleatoria.

Hasta aquí es todo muy bonito, pero surge otra pregunta muy importante: ¿existen datos reales que tengan distribuciones como las *de marras*? La verdad es que cuesta un poco encontrar, pero hay algunas que al menos se aproximan. Por ejemplo, el histograma de la Figura 2 corresponde a las estaturas de 137 adolescentes de 18 años [1], donde se observa una distribución bimodal aproximadamente simétrica. Esta forma es producto de las distintas alturas de hombres y mujeres presentes en la muestra. La estatura media y la mediana es de 1,67 y 1,66 metros respectivamente, aproximadamente en el punto medio de las modas 1,59 y 1,71 metros (considerando las marcas de clase). La asimetría de 0,216 es levemente positiva, lo que se alcanza a percibir en que la cola derecha es un poco más extendida que la izquierda.

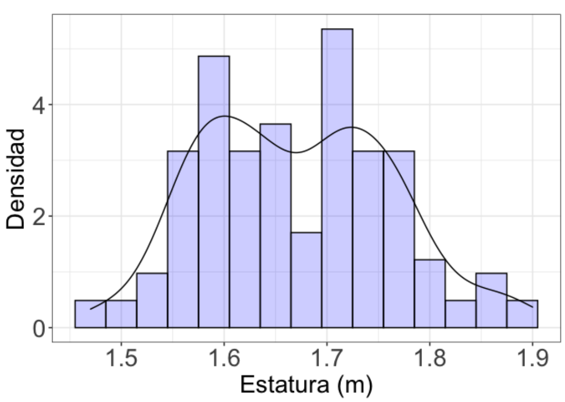

Figura 2: Histograma de estaturas de 137 adolescentes de 18 años y su función de densidad ajustada.

El histograma de la Figura 3 corresponde a datos de un estudio realizado durante uno de los períodos de cuarentena de la pandemia de COVID-19, a través de una encuesta telefónica. Específicamente, los datos son las respuestas dadas por una muestra de 477 mujeres del Programa de Salud Cardiovascular de la Atención Primaria a la siguiente pregunta: *¿En qué medida el equipo médico es un apoyo para usted en la pandemia?* La respuesta está expresada en un puntaje de 0 a 10 [2]. Aquí el puntaje promedio es 5,5 y la mediana es 6, también ambas cercanas al punto medio entre los valores extremos de 0 y 10. El coeficiente de asimetría es -0,22, es decir, levemente negativo, ya que en los extremos la moda menos frecuente (izquierda) produce el efecto equivalente a una cola más alargada en las distribuciones que son más habituales que acostumbramos ver.

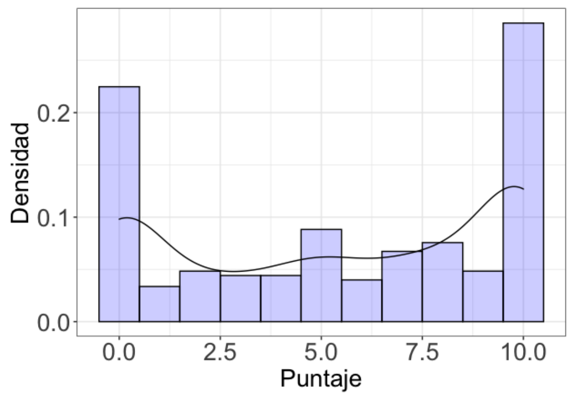

Figura 3: Histograma de respuestas a la pregunta *¿En qué medida el equipo médico es un apoyo para usted en la pandemia?*, junto a su función de densidad ajustada.

Entendiendo el concepto podemos imaginar infinitas distribuciones simétricas sin forma de campana y no necesariamente bimodales. Podemos partir desde la más simple y obvia, la uniforme, para luego pasar a multimodales de distinto tipo, algunas de las cuales se muestran en la Figura 4. Es evidente que encontrar datos con estas formas de distribución es mucho más difícil.

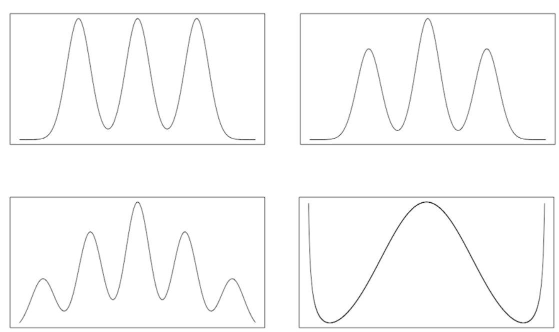

Figura 4: Otras distribuciones simétricas sin forma de campana de Gauss.

Como conclusión, si bien cuando la media, la mediana y la moda coinciden estamos frente a una distribución simétrica, lo opuesto no siempre es cierto. Esto quiere decir que si una distribución es simétrica no necesariamente la media, la mediana y la moda coinciden. Más aún, la simetría en una distribución no implica una forma de campana de Gauss, siendo esta última una de las tantas —en efecto, infinitas— formas simétricas posibles.

Tomando en cuenta el error que cometimos como equipo docente en la redacción de una de las alternativas, he aprovechado de escribir esta pequeña reflexión (en parte un *mea culpa*) en la que, además de la discusión estadística, quiero resaltar la importancia de que como profesores revisemos permanentemente lo que enseñamos, a la vez de incentivar la mirada crítica de parte de las y los estudiantes. Esto último es, en mi opinión, especialmente importante ya que debemos esforzarnos no sólo en formar personas con conocimientos específicos, sino también que sean capaces de poner en duda lo que se les transmite, tener la autonomía para buscar respuestas y discutir lo que les enseñamos.

Referencias

[1] Ministerio de Salud, Encuesta Mundial de Salud Escolar 2013, disponible en http://www.repositoriodigital.minsal.cl/handle/2015/908. Base de datos descargada desde http://epi.minsal.cl/bases-de-datos/?s=encuesta+mundial+de+salud+escolar.

[2] Nicoletti-Rojas, D., *et al*, Effects of sociodemographic and health factors on the self-management of non- communicable diseases among Chilean adults during the Covid-19 pandemic, *PLOS GLOBAL PUBLIC HEALTH*, 2022; *2(7)*: 1-16.