

La curtosis indica el peso de las colas, y nada más

MAURICIO FUENTES ALBURQUENQUE*

PROGRAMA DE BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA
FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE

1. Introducción

La distribución de una variable aleatoria, ya sea empírica o teórica, puede describirse mediante varias medidas de resumen, todas complementarias entre sí. Estas medidas entregan información de distintas características de la distribución: tendencia central, dispersión, posición y forma. Para esta última usamos típicamente dos medidas, a través de sus respectivos coeficientes: la asimetría y la curtosis.

La curtosis se ha interpretado tradicionalmente, al menos en algunas áreas de investigación, como el grado de lo «puntiagudo» de la distribución. Así, de acuerdo a la definición original introducida por Karl Pearson, el valor 3 correspondería a una distribución Gaussiana o normal (distribución «mesocúrtica») y, a partir de ahí, si es más puntiaguda es mayor a 3 (distribución «leptocúrtica») y si es menos puntiaguda es menor a 3 (distribución «platicúrtica»). Sin embargo, una mayor curtosis también indica la presencia de colas más pesadas, interpretación que no siempre está presente.

En este ensayo revisaremos esta medida, repasando su historia y la evolución de su interpretación. Concluiremos, o más bien redescubriremos la conclusión a la que algunos autores habían llegado hace tiempo, de que la curtosis nada tiene que ver con lo «puntiagudo» de la distribución sino solo con el peso de las colas.

2. Los orígenes

Cuenta la historia que fue Karl Pearson quien introdujo por primera vez el término *kurtosis* (en inglés), del griego *kurtos* (curvo). Según señalan varios autores y autoras, el término habría aparecido por primera vez en Pearson (1905), un artículo de réplica¹ a la crítica de Ranke & Greiner (1904) a uno de sus trabajos sobre las distribuciones asimétricas a las cuales no se ajusta la distribución normal.

En dicho artículo Pearson describe las que, a su juicio, son las tres principales diferencias entre las distribuciones de frecuencia reales (empíricas) y la distribución teórica Gaussiana o normal: (i) la separación entre la media y la moda; (ii) la razón entre esta separación y la variabilidad, a la que llamó *skewness* (que

conocemos en español como asimetría); y, cito textual (traducción propia),

(iii) Un grado de aplanamiento de la parte superior que es mayor o menor que el de la curva normal. Dadas dos distribuciones de frecuencia que tienen la misma variabilidad medida por la desviación estándar, pueden ser relativamente más o menos aplanadas que la curva normal. Si son más aplanadas, las denomino *platicúrticas*; si son menos aplanadas, *leptocúrticas*; y si son igualmente aplanadas, *mesocúrticas*. Una distribución de frecuencia puede ser simétrica, satisfaciendo las dos primeras condiciones de normalidad, pero puede no ser *mesocúrtica*, y por lo tanto la curva Gaussiana no puede describirla.

Del escrito queda la impresión que Pearson ya hubiese mencionado o definido antes el concepto de curtosis. Sin embargo, aunque en diversos artículos desde 1894 venía desarrollando el concepto (Fiori & Zenga, 2009), al parecer esto no fue así, lo que explicaría que varios autores y autoras señalen el artículo de 1905 como aquel en el que se introdujo el término (Balanda & MacGillivray, 1988; Dyson, 1943; Fiori & Zenga, 2009; Westfall, 2014).

Pearson había definido antes lo que podríamos traducir como «aplanamiento de la cima» o «achatación» (*flat-toppedness*) (Pearson, 1902), dado por

$$\beta_2 = \frac{\mu_4}{\sigma^4}, \quad (1)$$

donde $\mu_4 = E[(X - E[X])^4]$ es el cuarto momento de la distribución en torno a la media y $\sigma^2 = E[(X - E[X])^2]$ es la varianza (segundo momento en torno a la media) de la variable aleatoria X . La expresión (1) se conoce también como el cuarto momento estandarizado y es lo que generalmente se define como curtosis, que para la curva Gaussiana toma el valor 3. Pearson (1905) llamó a la cantidad $\beta_2 - 3$ el «grado de curtosis» (*degree of kurtosis*), cuyo valor para la distribución normal es, naturalmente, cero. Comúnmente se define también $\gamma_2 = \beta_2 - 3$ como el «exceso de curtosis».

* mauricio.fuentes@uchile.cl.

¹ Esta réplica tiene la nada despreciable extensión de 44 páginas, algo usual en los artículos o «memorias» (*memoirs*) que escribía Pearson. Quizás sus estudios de Leyes (Magnello, 2005) hayan influido algo en su inclinación a dar extensos argumentos de sus ideas.

Para una completa revisión del desarrollo del concepto de curtosis por parte de Karl Pearson, recomiendo el trabajo de Fiori & Zenga (2009).

3. Las interpretaciones

Karl Pearson, en distintos artículos, destacó la relación de la curtosis solo con el grado de achatamiento, sin darle aparentemente otra interpretación. Sin embargo, algunos años después otros autores comienzan a señalar que la curtosis también se relaciona con lo pesado de las colas. Ronald Fisher, por ejemplo, señala que la estadística calculada a partir del cuarto momento es «una medida de desviación de la normalidad, en este caso de tipo simétrico, mediante la cual el ápice y las dos colas de la curva aumentan a expensas de la porción intermedia, o cuando el correspondiente parámetro $\gamma_2 (= \beta_2 - 3)$ es negativo, la cima y las colas se reducen y los hombros se rellenan, formando una curva relativamente plana en la parte superior» (Fisher, 1934). *Student*, citado por Yule & Kendall (1937), dio una «mnemotecnia divertida» señalando que «Las curvas platicúrticas, como el ornitorrinco [juego de palabras con el término en inglés *platypus*], son rechonchas con colas cortas. Las curvas leptocúrticas son altas con colas largas como el canguro –reconocido por “saltarín”! [*lepping*]». Dyson (1943) lo hace implícitamente al establecer la condición suficiente para que dos funciones de densidad tengan distintos valores de curtosis. Finucan (1964), posteriormente, afirma que su artículo «“redescubre” la interpretación original de la curtosis como un indicador de “un prominente pico y una prominente cola” en la curva de densidad».

Vemos que Pearson, siendo quien propuso la curtosis, no dio una interpretación correcta y solo se refirió al aplanamiento (o a la «puntiagudez»²) de la parte superior de la curva de densidad. Esto habría sido en parte corregido y ya aparece en los escritos de Fisher, agregando el peso de las colas. Sin embargo, como señalan Dyson (1943) y Kaplansky (1945), algunos textos de Estadística como Rietz (1927), Yule & Kendall (1937), Kenney (1939) y Aitken (1939) replicaron la interpretación exclusiva del aplanamiento de Pearson. Podemos agregar a esta lista a Snedecor (1946), D’Ottone (1965), García Hoz & Ferrer Martín (1966), Snedecor & Cochran (1967), Cortada de Kohan & Carrero (1970), Mood et al. (1974), Canavos (1988), Spiegel (1991), Norman & Streiner (1998), Espina Marconi (1998), Casella & Berger (2002), van Belle et al. (2004) y Martínez González et al. (2014). Westfall (2014) nombra y cita textual extractos de otros libros, como Sapp (2006), Reinard (2006), McDonald (2007), Taylor (2008), Lee et al. (2013), Coolican (2013) y

Katz et al. (2013), además de varios artículos de revistas con revisión de pares.

También hay varios autores y autoras que dan la interpretación mixta, como la de Fisher. De acuerdo a Balanda & MacGillivray (1988), la característica de forma que llamamos curtosis «puede ser vagamente definida como el movimiento, independiente de la posición y la escala, de masa de probabilidad desde los hombros de una distribución hacia el centro y las colas». Esta interpretación es explicada con más detalle por DeCarlo (1997), quien señala (traducción propia):

...la curtosis representa un movimiento de masa que no afecta la varianza. Considérese el caso de curtosis positiva, donde las colas más pesadas son acompañadas a menudo por un pico más alto. Nótese que si la masa se mueve simplemente desde los hombros de una distribución hacia sus colas, entonces la varianza también será mayor. Para conservar la varianza uno debe mover también masa desde los hombros hasta el centro, lo que compensará disminuyendo la varianza y dará un pico mayor. Para la curtosis negativa, la varianza se conservará si la masa se mueve desde las colas y el centro de la distribución hacia sus hombros, resultando en colas livianas y aplanamiento... La explicación anterior divide la distribución en colas, hombros y centro, donde para β_2 los hombros están ubicados en $\mu \pm \sigma$...

DeCarlo (1997) ilustra lo anterior comparando una familia de distribuciones beta simétricas estandarizadas y platicúrticas, con distintos valores para el parámetro de forma, con la distribución normal estándar. En la Figura 1 se muestran las gráficas de estas distribuciones, en donde se observa el ensanche de los hombros y la reducción de las colas y el pico a medida que decrece la curtosis (es decir, se hace más negativa). El caso más extremo es el bimodal, donde γ_2 alcanza un valor considerablemente negativo³. En el artículo DeCarlo discute que la negatividad de la curtosis no se relaciona necesariamente con la bimodalidad, como habría planteado Darlington (1970), citando contraejemplos de Hildebrand (1971) y Balanda & MacGillivray (1988).

Es necesario aclarar que para interpretar la curtosis de una distribución estamos considerando como base de comparación una distribución gaussiana con igual varianza. En otras palabras, la curtosis indicará si nuestra distribución tiene colas más pesadas y/o es más puntiaguda en el centro⁴ comparada con una distri-

²Aunque, de acuerdo al diccionario de la Real Academia Española (<https://dle.rae.es>), este término no existe en español, lo usaré aquí como una traducción del término *peakedness*, usado comúnmente en inglés en el contexto de la curtosis.

³El valor mínimo que puede tomar β_2 es 1 (ver nota 10), por lo que el mínimo de γ_2 es -2.

⁴Por ahora lo planteo así, aunque la conclusión de este ensayo es que la curtosis es una medida solo del peso de las colas.

bución normal con la misma varianza que *nuestra* distribución. También estamos aquí tomando en cuenta que el grado de curtosis lo vemos de manera independiente de la asimetría de la curva. Todo esto es consistente con la interpretación original de Pearson. Para discusiones sobre la relación entre simetría y curtosis y otras propuestas para medir esta última, se puede revisar, por ejemplo, Balanda & MacGillivray (1988), Royston (1992) y DeCarlo (1997).

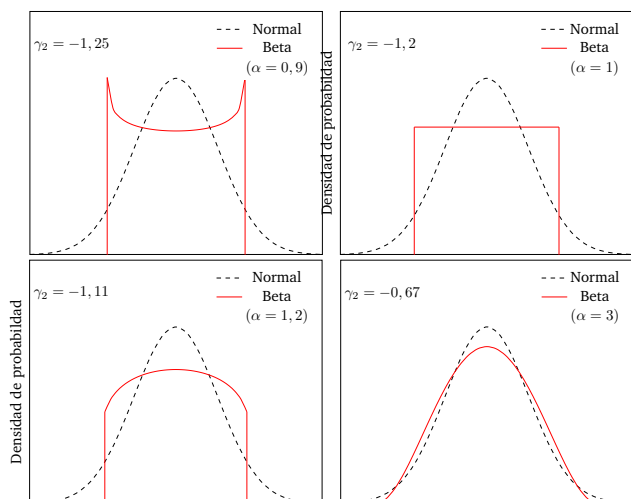


Figura 1. Comparación de distribuciones beta simétricas estandarizadas, con parámetro de forma α y diferentes curtosis, con la distribución normal estándar (modificadas de DeCarlo (1997), Figura 5).

Con todo lo anterior, es difícil saber cuál de las dos interpretaciones, la *Pearsoniana* o la *mixta* (*Pearsoniana* combinada con las colas), ha sido la más frecuente. Lo que sí es bastante claro es que la interpretación exclusiva del peso de las colas ha sido minoritaria. A continuación describiremos dos publicaciones que, podríamos decir, barrieron con la versión del aplanamiento.

4. Kaplansky (1945): el golpe mortal a la interpretación pearsoniana

En un artículo de tan solo una página, Kaplansky (1945) entrega cuatro funciones de densidad, todas simétricas en torno al origen, con media 0 y varianza 1, con diferentes valores máximos en $x = 0$ y distintos valores de curtosis⁵. Estos últimos solo dependen del cuarto momento central μ_4 , ya que como la varianza es 1, la curtosis es $\beta_2 = \mu_4/\sigma^4 = \mu_4$. Las funciones dadas por Kaplansky (1945) se muestran en la Tabla 1⁶.

En el artículo, así se expresa el problema y su objetivo (traducción propia):

En muchos textos se establece que una curva de frecuencia con curtosis positiva es mayor en el entorno de la media que la correspondiente curva normal, mientras que una con curtosis negativa es menor. Con la esperanza de aclarar este error, ofrecemos en esta nota cuatro ejemplos mostrando que puede ocurrir cualquier combinación de puntiagudez [*peakedness*] en la media y curtosis.

En efecto, de la Tabla 1 vemos que las funciones f_1 y f_2 son «platicúrticas» ($\mu_4 < 3$), pero $f_1(0)$ es menor y $f_2(0)$ es mayor que la densidad gaussiana en $x = 0$, que toma aproximadamente el valor 0,399. Lo mismo ocurre con f_3 y f_4 , ambas «leptocúrticas» ($\mu_4 > 3$). Esto se muestra en las gráficas de la Figura 2.

No cabe duda que Kaplansky logró su objetivo y refutó, en un sentido «Popperiano» podríamos decir, la interpretación Pearsoniana. Ahí debió morir tal interpretación, pero de alguna manera sobrevivió. Quizás, creo, el excesivo laconismo del artículo (o nota) haya influido. Posiblemente, aunque mantuviera el carácter breve de su nota, si hubiese agregado las gráficas de las funciones hubiese quedado claro el punto, ya que salta a la vista de inmediato (ver Figura 2), y su impacto habría sido más efectivo. No lo podemos saber.

Considerando que Kaplansky se enfocó solo en las puntas o máximos de las distribuciones, podríamos inferir que era parte de quienes no sabían de la interpretación de las colas. Sin embargo, si miramos con aumento hacia los extremos de las distribuciones, como se muestra en la Figura 2, podemos notar que las funciones «platicúrticas» f_1 y f_2 tienen colas menos pesadas que la distribución gaussiana, mientras que las «leptocúrticas» f_3 y f_4 presentan la característica contraria. Si bien no podemos generalizar esto a partir de cuatro casos, sí tenemos un valioso ejemplo en que la curtosis se relaciona solo con lo pesado de las colas. Seguramente Kaplansky no lo vio ya que no era lo que pretendía encontrar, y es posible que esto no hubiese sido distinto aún incluyendo las gráficas, lo que daría la razón a lo afirmado por Whitehead (1925) (citado por Taleb (2008)):

Llegar muy cerca de la teoría correcta y captar su aplicación exacta son dos cosas muy distintas, como bien nos enseña la historia de la ciencia. En todo lo importante, siempre ha habido alguien que lo ha dicho antes pero no lo ha descubierto.

⁵Los momentos, el valor en el origen y que la integración en el dominio de los números reales sea 1 lo verifiqué usando R.

⁶Les he cambiado los nombres respecto a cómo las llamó Kaplansky (1945), y las ordené según sus valores de curtosis para facilitar el análisis.

Tabla 1: Funciones de densidad† dadas por Kaplansky (1945).

Función	Valor en $x = 0$ (máximo)	Curtosis (β_2)
$f_1(x) = \frac{3\sqrt{3}}{16\sqrt{\pi}}(2 + x^2)e^{-3x^2/4}$	0,367	2,67
$f_2(x) = \frac{1}{3\sqrt{\pi}}\left(\frac{9}{4} + x^4\right)e^{-x^2}$	0,423	2,75
$f_3(x) = \frac{3}{2\sqrt{2\pi}}e^{-x^2/2} - \frac{1}{6\sqrt{\pi}}\left(\frac{9}{4} + x^4\right)e^{-x^2}$	0,387	3,13
$f_4(x) = \frac{1}{6\sqrt{\pi}}(e^{-x^2/4} + 4e^{-x^2})$	0,470	4,50

† Todas con primeros momentos $\mu = 0$, $\sigma^2 = 1$ y $\mu_3 = 0$.

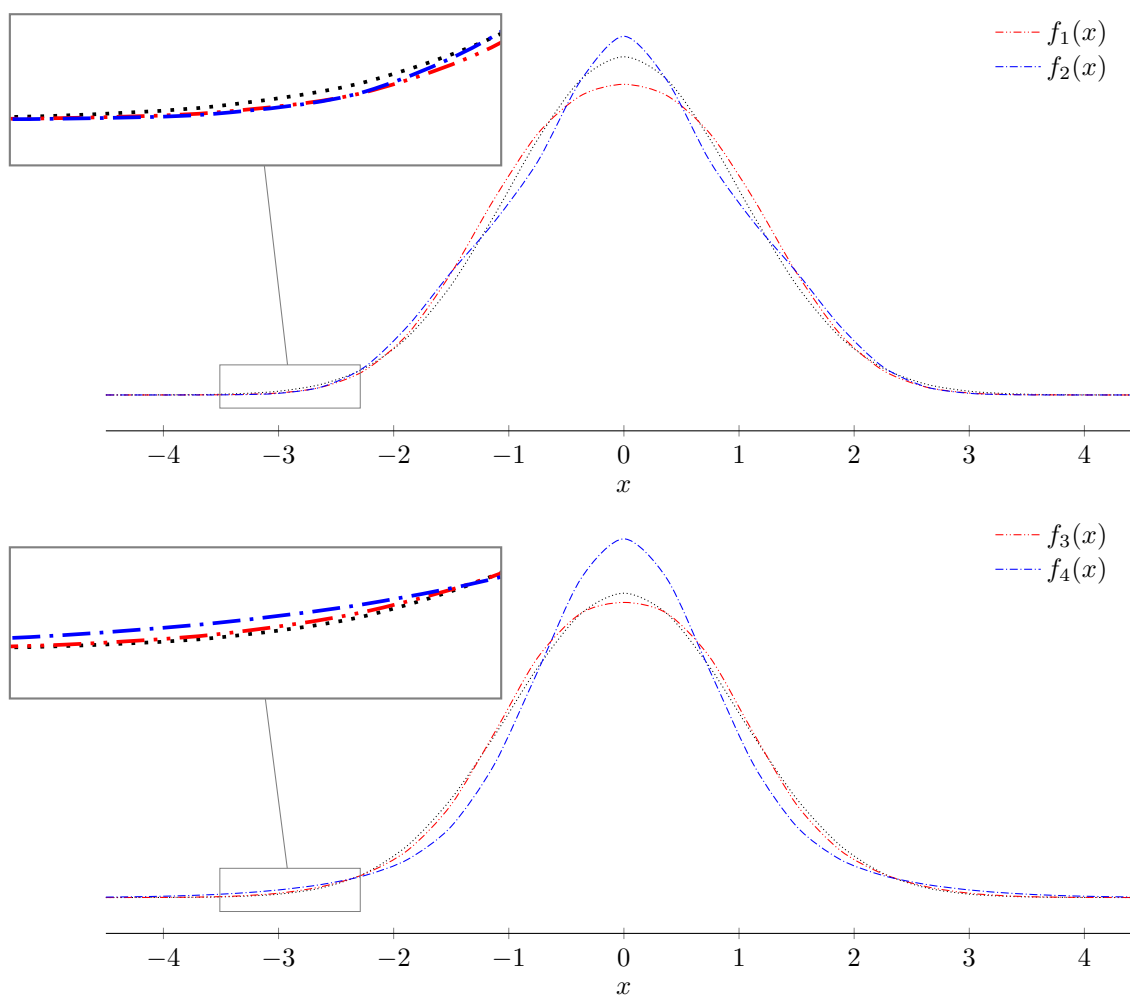


Figura 2. Funciones de densidad dadas por Kaplansky (1945) y la densidad Gaussiana (línea punteada). En el gráfico superior se muestran las funciones «platicúrticas» f_1 y f_2 . En el gráfico inferior se muestran las funciones «leptocúrticas» f_3 y f_4 .

5. Westfall (2014): el tiro de gracia

Con el sugerente título *Kurtosis as Peakedness, 1905-2014. R.I.P.*, Westfall (2014) comienza repasando la gran cantidad de fuentes bibliográficas, tanto libros como artículos de prestigiosas revistas, que continuaban a la fecha reproduciendo las interpretaciones Pearsoniana y mixta (ya he mencionado los textos en la sección 3).

Westfall señala que «Quizás la persistencia del término “puntiagudez” [*peakedness*] viene de personas que miran los histogramas de datos con colas pesadas y ven un “pico” [*peak*] fuertemente pronunciado». Para apoyar esta impresión usa una muestra aleatoria de 1.000 datos simulados desde una distribución Cauchy⁷, cuyo histograma se muestra en la Figura 3. Agrega que «El gráfico parece mostrar un visible “pico” angosto en el centro, pero esto es solo un artefacto de la escala del eje x . Son los *outliers*, esto es, las colas, lo que determina su apariencia».

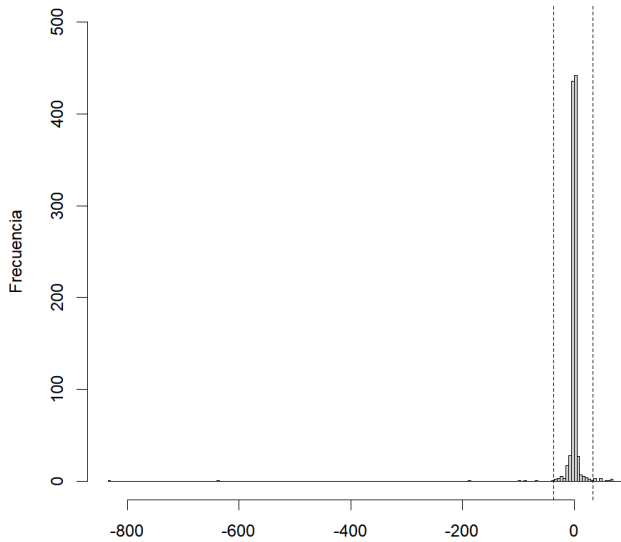


Figura 3. Histograma de una muestra aleatoria Cauchy de tamaño 1.000. Las líneas punteadas corresponden al intervalo $\bar{x} \pm \sigma$ (modificado de Westfall (2014), Figura 1).

Luego, considerando la media y la desviación estándar empíricas, $\bar{x} = -1,55$ y $s = 34,85$ ⁸, respectivamente, construye la variable $z_i = (x_i - \bar{x})/s$. Entonces, la curtosis empírica es

$$\beta_2 = \frac{1}{1000} \sum_i z_i^4 = 437,3871. \quad (2)$$

Separando la curtosis en la porción dentro y fuera

⁷Declara hacerlo con la función `rcauchy` de R con la semilla 12344, lo que pude verificar.

⁸«usando n en el denominador para que corresponda a la distribución empírica» (Westfall, 2014).

⁹ $E[Z^4] = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] = \frac{E[(X - \mu)^4]}{\sigma^4} = \frac{\mu_4}{\sigma^4} = \beta_2$.

de una desviación estándar tenemos

$$\begin{aligned} \beta_2 &= \frac{1}{1000} \sum_{|z_i| \leq 1} z_i^4 + \frac{1}{1000} \sum_{|z_i| > 1} z_i^4 \\ &= 0,0073 + 437,3798 = 437,3871. \end{aligned} \quad (3)$$

De (3) se desprende que, para esta distribución, la porción de curtosis determinada por la parte central es $0,0073/437,3871 = 0,000017$, es decir, ¡17 millonésimas!

En la segunda parte de su argumentación, Westfall usa tres distribuciones simétricas, llamadas «torre del diablo» (*devil's tower*, plana en la porción central), «triangular» (puntuda en el centro) y «vestido lencerro» (*slip-dress*, bimodal con picos infinitos). Muestra un gráfico con estas tres distribuciones, que tienen todas curtosis $\beta_2 = 2, 4$, y luego otro gráfico con las mismas tres con $\beta_2 \approx 6000$. Ambos gráficos lucen idénticos. Con esto el autor justifica el título de la sección: «Ni una curtosis grande ni una pequeña entregan información útil acerca de la forma de la punta».

Antes de entregar su argumento más general, el autor señala que «la noción de que κ [β_2 para nosotros] es informativo sobre el pico de una distribución es sospechosa: los valores esperados son promedios, y los promedios están fuertemente influenciados por los extremos». Esta reflexión es totalmente razonable y parece demasiado obvia, por lo que llama la atención que tan pocos lo hayan notado en más de un siglo, descontando, por cierto, al propio Pearson.

Siguiendo a Westfall (2014), supongamos que nuestro interés se centra en la variable aleatoria X que sigue una distribución con media μ , varianza σ^2 , cuarto momento μ_4 y curtosis β_2 . Definamos $Z = (X - \mu)/\sigma$. Es claro que $E[Z^4] = \beta_2$ ⁹. Sea h una cantidad positiva, F_Z la distribución de Z y $A_h = \{z : |z| \leq h\}$ una porción central de F_Z . Siguiendo la idea de (3), la curtosis la podemos particionar como

$$\begin{aligned} \beta_2 &= \int_Z z^4 dF_Z = \int_{A_h} z^4 dF_Z + \int_{A_h^c} z^4 dF_Z \\ &= C_h + T_h, \end{aligned} \quad (4)$$

donde A_h^c es el área *no central* (la unión de las áreas correspondientes a las colas, el complemento de A_h), $C_h = \int_{A_h} z^4 dF_Z$ es la porción central de β_2 y $T_h = \int_{A_h^c} z^4 dF_Z$ es la porción *no central* o de las colas de β_2 .

Para la porción central tenemos

$$\begin{aligned} |z| &\leq h \\ z^4 &\leq h^4 \\ \int_{-h}^h z^4 dF_Z &\leq \int_{-h}^h h^4 dF_Z = h^4 \underbrace{\int_{-h}^h dF_Z}_{\leq 1} \leq h^4, \end{aligned}$$

es decir,

$$0 \leq C_h \leq h^4,$$

y sumando T_h tenemos

$$T_h \leq \beta_2 \leq T_h + h^4. \quad (5)$$

Westfall demuestra que si la función de densidad de Z^2 es decreciente en el intervalo $[0, 1]$, entonces el valor esperado de Z^4 en la porción central está acotado superiormente por $1/2$. La misma cota se cumple si la densidad de Z es simétrica y se cumple que $f_Z(z)/z$ es decreciente para $0 < z < 1$.

En forma general, si miramos con detención la desigualdad (5), vemos que si la porción de las colas crece mucho, las cotas inferior y superior de β_2 van a estar dominadas por T_h . Esto es más claro aún si pensamos que h no puede ser demasiado grande (de modo que represente efectivamente una zona *central* de la distribución y no se extienda hacia la zona de las colas). Particularmente, para Z el valor de h debería ser 1 si consideramos la definición dada comúnmente para los «hombros» de la distribución (como lo hace DeCarlo (1997)), aunque podríamos perfectamente definirlo con un valor menor. En cualquier caso, una cota razonable es $h \leq 1$, es decir, $h^4 \leq 1$. Luego

$$T_h \leq \beta_2 \leq T_h + 1. \quad (6)$$

De (6) tenemos que si la porción de las colas tiende a cero, β_2 quedaría acotada entre 0 y 1, lo que equivale a que $\beta_2 = 1$ (su valor mínimo¹⁰). Con todo lo dicho, queda claro que las colas explican una proporción mucho mayor de la curtosis que el centro.

Lo anterior, evidentemente, no es una demostración formal. Tampoco Westfall (2014) la entrega (salvo para los casos particulares mencionados), pero la inspección y el análisis que hace de estas desigualdades, y que he intentado transmitir aquí, es bastante sólido, en mi opinión, como para convencernos de su argumento.

Para finalizar y otorgar evidencia más tangible, el artículo de Westfall entrega la proporción de curtosis determinada por el centro ($\int_{\mu \pm \sigma} z^4 dF_Z$), para varias distribuciones de probabilidad conocidas. Las distribuciones están ordenadas desde la proporción central mayor a la menor. Solo algunas especiales, como Bernoulli y normales mezcladas, presentan una proporción central grande o importante en la curtosis, pero en la gran mayoría esta proporción no llega al 10%. Parece bastante razonable, entonces, estar de acuerdo con la conclusión de Westfall (2014) (traducción propia):

... la curtosis te dice muy poco sobre el pico o centro de la distribución. De este modo, la curtosis no debería definirse nunca en términos de la puntigudez [*peakedness*].

Hacerlo es contraproducente con el propósito de fomentar la alfabetización estadística. La relación de la puntigudez con la curtosis se acaba oficialmente ahora.

Tabla 2: Proporción de curtosis determinada por la parte central $\mu \pm \sigma$ para varias distribuciones de probabilidad (Westfall, 2014).

Distribución	Curtosis (β_2)	Centro
Ber(1/2)	1,00	1,000
Ber(1/2 + ϵ)	1,00	0,500
$\frac{1}{2}N(-9, 1) + \frac{1}{2}N(9, 1)$	1,05	0,352
$\frac{1}{2}N(-2, 1) + \frac{1}{2}N(2, 1)$	1,72	0,099
Uniforme	1,80	0,064
Vestido lencero	2,40	0,063
Torre del diablo	2,40	0,055
$\frac{1}{2}N(-1, 1) + \frac{1}{2}N(1, 1)$	2,50	0,048
Triangular simétrica	2,40	0,045
Normal	3,00	0,037
$t(10)$	4,00	0,028
Logística	4,20	0,026
Exponencial	9,00	0,023
Laplace	6,00	0,015
$t(5)$	9,00	0,012
Empírica (Figura 3)	437,4	$1,7 \times 10^{-5}$

6. Reflexiones finales y conclusión

La curtosis, al igual que la asimetría, nació como una propuesta de Karl Pearson para evaluar si la distribución teórica Gaussiana se ajusta a una distribución empírica dada. Este problema fue trabajado también por otros autores (Hald, 2004), aunque fue Pearson quien profundizó más en él, motivado justamente por los hallazgos en datos reales. Se enfrentó, muy en su estilo, a otros investigadores que seguían la doctrina de la distribución normal como modelo probabilístico para muchas (si no todas) las distribuciones encontradas en la naturaleza.

Sin embargo, en esta historia ocurrió un hecho poco usual: el propio creador de la medida dio una interpretación errónea de la misma, lo que se propagó y

¹⁰Sea la función $u(X) = X^2$. Dado que u es convexa se cumple la desigualdad de Jensen: $E[u(X)] \geq u(E[X])$ (Resnick, 2005). Ahora, sea $X = Z^2$, luego $E[Z^4] \geq (E[Z^2])^2 = 1$, es decir, $\beta_2 \geq 1$.

se extendió en el tiempo (probablemente influenciado por la autoridad que representaba Pearson en la Biometría y la Estadística). Esta interpretación errónea consiste en señalar que la curtosis es una medida de la «puntiagudez» (*peakedness*) de la distribución. Algunos años más tarde surgió una segunda interpretación que se mezcló con la original: que la curtosis mide lo puntiagudo y el peso de las colas de la distribución. Si bien esta interpretación «mixta» aparece señalada por Fisher tempranamente, no sabemos si fue el primero en hacerla pública.

Según lo expuesto en este ensayo, a mi juicio hay dos hitos relevantes y decisivos que derriban la interpretación Pearsoniana de la puntiagudez y favorecen la interpretación de las colas. El primero es el artículo de Kaplansky (1945), que entrega ejemplos de funciones de densidad donde ya sea la curtosis en exceso (leptocurtosis) o en defecto (platicurtosis) pueden ocurrir independientemente de si la densidad al centro de su distribución está por encima o por debajo del de la curva Gaussiana. En mi opinión, este artículo refuta inapelablemente la interpretación Pearsoniana al entregar casos particulares que la contradicen. Por lo mismo, es llamativo que varios autores que han escrito sobre este tema citen a Kaplansky (1945) y luego hayan seguido incluyendo la parte central de la distribución dentro de la interpretación, por ejemplo Finucan (1964), Balanda & MacGillivray (1988) y DeCarlo (1997).

El segundo hito es el artículo de Westfall (2014), en donde muestra en forma general que las colas son las que más determinan el valor de la curtosis, mientras que el centro de la distribución tiene un papel secundario y cada vez menor, llegando a constituir una fracción despreciable, a medida que la curtosis crece. En otras palabras, la magnitud de la curtosis en realidad estaría relacionada con la extensión de las colas, desde su valor mínimo 1 (ausencia de colas), pasando por el 3 (colas idénticas a la curva gaussiana) y extendiéndose indefinidamente (colas muy pesadas).

Lo expuesto implica, en términos prácticos, interpretar una mayor curtosis como una señal de mayor propensión a observar valores extremos, es decir, observaciones alejadas del centro de la distribución, con posibles consecuencias para la detección de anomalías y la elección de métodos estadísticos robustos.

De acuerdo a todo lo anterior, es necesario que eliminemos de nuestro lenguaje estadístico, tanto en investigación como en docencia, el concepto de curtosis como una medida de lo puntiagudo de una distribución y la usemos, y enseñemos, como una medida exclusivamente del grueso o peso de las colas. Es un error que se ha prolongado por demasiado tiempo y que debemos corregir.

Agradecimientos

Quiero agradecer a mis colegas Pía Muñoz Quezada y Felipe Medina Marín por sus valiosos comentarios y sugerencias (el penúltimo párrafo es una sugerencia casi textual del profesor Medina). Estos permitieron mejorar el escrito y lograr esta versión final (nunca perfecta, pero mejor que la original y la mejor que se pudo).

Referencias

- Aitken, A. C. (1939). *Statistical Mathematics*. Oliver and Boyd.
- Balanda, K. P., & MacGillivray, H. L. (1988). Kurtosis: A Critical Review. *The American Statistician*, 42(2), 111-119. <https://doi.org/10.2307/2684482>
- Canavos, G. C. (1988). *Probabilidad y Estadística. Aplicaciones y métodos*. McGraw-Hill.
- Casella, G., & Berger, R. L. (2002). *Statistical Inference* (2nd edition). Cengage Learning.
- Coolican, H. (2013). *Research Methods and Statistics in Psychology* (5th edition). Routledge.
- Cortada de Kohan, N., & Carro, J. M. (1970). *Estadística aplicada* (4a edición). Editorial Universitaria de Buenos Aires.
- Darlington, R. B. (1970). Is Kurtosis Really “Peakedness?” *The American Statistician*, 24(2), 19-22. <https://doi.org/10.2307/2681925>
- DeCarlo, L. T. (1997). On the Meaning and Use of Kurtosis. *Psychological Methods*, 2(3), 292-307. <https://doi.org/10.1037/1082-989X.2.3.292>
- D’Ottone, H. (1965). *Elementos de Estadística básica* (2a edición). Cooperativa de Cultura y Publicaciones, Ltda.
- Dyson, F. J. (1943). A Note on Kurtosis. *Journal of the Royal Statistical Society*, 106(4), 360-361. <https://doi.org/10.1111/j.2397-2335.1943.tb01574.x>
- Espina Marconi, L. (1998). *estadística Elemental* (6a edición). Espina Marconi.
- Finucan, H. M. (1964). A Note on Kurtosis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26, 111-112. <https://doi.org/10.1111/j.2517-6161.1964.tb00545.x>
- Fiori, A. M., & Zenga, M. (2009). Karl Pearson and the Origin of Kurtosis. *International Statistical Review*, 77(1), 40-50. <https://doi.org/10.1111/j.1751-5823.2009.00076.x>
- Fisher, R. A. (1934). *Statistical Methods for Research Workers* (5th edition). Oliver and Boyd.
- García Hoz, V., & Ferrer Martín, S. (1966). *Estadística Aplicada a la Educación y Ciencias Humanas*. Ediciones RIALP.

- Hald, A. (2004). *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713 to 1935*. University of Copenhagen.
- Hildebrand, D. K. (1971). Kurtosis measures bimodality? *American Statistician*, 25(1), 42-43. <https://doi.org/10.1080/00031305.1971.10477241>
- Kaplansky, I. (1945). A Common Error concerning Kurtosis. *Journal of the American Statistical Association*, 40(230), 259-259. <https://doi.org/10.1080/01621459.1945.10501856>
- Katz, D. L., Elmore, J. G., Wild, D. M., & Lucan, S. C. (2013). *Jekel's Epidemiology, Biostatistics and Preventive Medicine* (4th edition). Elsevier.
- Kenney, J. F. (1939). *Mathematics of Statistics*. D. Van Nostrand Company, Inc.
- Lee, C. F., Lee, J. C., & Lee, A. C. (2013). *Statistics for Business and Financial Economics* (3rd edition). Springer.
- Magnello, M. E. (2005). Pearson, Karl. En P. Armitage & T. Colton (Eds.), *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd.
- Martínez González, M. Á., Sánchez-Villegas, A., Toledo Atucha, E. A., & Faulin Fajardo, J. (2014). *Bioestadística amigable* (3a edición). Elsevier.
- McDonald, T. H. (2007). *Basic Concepts in Statistics and Epidemiology*. Radcliffe.
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3rd edition). McGraw-Hill.
- Norman, G. R., & Streiner, D. L. (1998). *Biostatistics. The Bare Essentials*. B.C. Decker Inc.
- Pearson, K. (1902). On the Mathematical Theory of Errors of Judgement, with Special Reference to the Personal Equation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 198, 235-299. <https://doi.org/10.1098/rsta.1902.0005>
- Pearson, K. (1905). "Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson." A Rejoinder. *Biometrika*, 4(1/2), 169-212. <https://doi.org/10.2307/2331536>
- Ranke, K. E., & Greiner. (1904). Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson. *Archiv für Anthropologie*, II, 295-331.
- Reinard, J. C. (2006). *Communication Research Statistics*. Sage.
- Resnick, S. I. (2005). *A Probability Path*. Birkhäuser.
- Rietz, H. L. (1927). *Mathematical Statistics*. The Mathematical Association of America.
- Royston, P. (1992). Which Measures of Skewness and Kurtosis are Best? *Statistics in Medicine*, 11, 333-343. <https://doi.org/10.1002/sim.4780110306>
- Sapp, M. (2006). *Basic Psychological Measurement, Research Designs, and Statistics Without Math*. Charles C. Thomas.
- Snedecor, G. W. (1946). *Métodos de Estadística: su aplicación a experimentos en Agricultura y Biología* (4a edición). Acme Agency.
- Snedecor, G. W., & Cochran, W. G. (1967). *Statistical Methods* (6th edition). The Iowa State University Press.
- Spiegel, M. R. (1991). *Estadística* (2a edición). McGraw-Hill.
- Taleb, N. N. (2008). *El cisne negro. El impacto de lo altamente probable*. Paidós.
- Taylor, S. E. (2008). Kurtosis. En S. Boslaugh & L.-A. McNutt (Eds.), *Encyclopedia of Epidemiology*. Sage Publications.
- van Belle, G., Fisher, L. D., Heagerty, P. J., & Lumley, T. (2004). *Biostatistics. A Methodology for the Health Sciences* (2nd edition). John Wiley & Sons, Inc.
- Westfall, P. H. (2014). Kurtosis as Peakedness, 1905–2014. R.I.P. *The American Statistician*, 68(3), 191-195. <https://doi.org/10.1080/00031305.2014.917055>
- Whitehead, A. N. (1925). *Science and the Modern World*. MacMillan Company.
- Yule, G. U., & Kendall, M. G. (1937). *An Introduction to the Theory of Statistics* (11th edition). Charles Griffin & Company, Ltd.