

## La “distribución normal” de Gauss y el “hombre tipo” de Quetelet

*De ti se trata en esta fábula*  
Horacio (*Sátiras*)

### La distribución normal de Gauss

Estudiando las variaciones que se producen al medir repetidamente un mismo objeto (largo de una mesa, posición de un planeta), o sea lo que se llamaría “distribución de los errores de medición”, Gauss describió una ecuación cuyo gráfico es la famosa “curva normal” o “curva de Gauss”, (1) que es de forma acampanada y simétrica, cuya altura y ancho dependen del valor de la varianza de la población; cuanto mayor sea la varianza (mayor es la dispersión) y más baja y ancha es la curva. La curva de Gauss estándar tiene una media del universo ( $\mu$ ) = 0, y una varianza ( $\sigma^2$ ) = 1 (Figura 1).

A esa “distribución de los errores de medición” Gauss la llamó “distribución normal”, en el sentido de “norma”, porque al decir “distribución normal de los errores de medición”, quería decir que ésa era la “norma de distribución de los errores de medición” de un solo y mismo objeto. Entonces, al hablar del límite fiduciario (del latín *fides* = confianza) intervalo de confianza del 95%, se quiere decir que entre esos límites tendríamos el 95% de los errores posibles de medición. A Gauss nunca se le ocurrió pensar que su ecuación y gráfico se utilizaría para hablar de “distribución normal” de un grupo o serie de valores individuales distintos. Podemos reconocer esta diferencia crucial utilizando un ejemplo muy simple.

Si estamos en la situación en que a 1 voluntario le extraemos 1 litro de sangre de la que separamos 500 ml de plasma, y sobre estos 500 ml hacemos 1.000 determinaciones de triglicéridos. En un mundo de

perfección infinita (ideal) los 1.000 resultados deberían ser exactamente iguales o idénticos, pero en la vida real no lo son debido a que las mediciones tienen cierto error. Estos 1.000 datos por lo tanto tendrán una “distribución normal de los errores de medición” o sea una “curva de distribución gaussiana”.

Si ahora en lugar de hacer las 1.000 determinaciones en un único voluntario, tomamos 1.000 voluntarios y le efectuamos a cada uno una determinación de triglicéridos plasmáticos, la distribución de estos 1.000 datos (que ahora van a corresponder a una población de 1.000 individuos y no a 1.000 mediciones del mismo individuo) no siguen la distribución descrita por Gauss. (2)

En realidad, la curva de distribución de los triglicéridos en una población de individuos es completamente asimétrica, con una concentración mayor dentro de los valores más bajos (considerados clínicamente normales) hacia la izquierda del gráfico y una pendiente larga y prolongada hacia los valores más altos; a esta distribución en los valores de una población se la llama con “moda a la izquierda” (ya que la moda es el valor que se repite más veces o de mayor frecuencia). La distribución de las variables biológicas pueden tener muchas formas comprobables empíricamente; la TSH también tiene una distribución asimétrica con moda a la izquierda, la altura de la población de un colegio va a mostrar dos picos de frecuencia o modas, primero la de los alumnos y luego la de los maestros, por eso se la llama distribución “bimodal”, y hay otras distintas.

Como corolario podemos decir si bien los “errores de medición” se distribuyen como lo describió Gauss, **los datos individuales de las variables biológicas no deben necesariamente distribuirse según la ecuación de Gauss.** Esto permite dos simples conclusiones con implicaciones clínicas importantes.

La primera es que si la distribución no es gaussiana, la “media” ya no describe la mejor medida de “tendencia central”, ni en el “intervalo de confianza del 95%” (IC 95%) se encontrarán 95 de cada 100 mediciones. Por lo tanto, deberán utilizarse otras medidas de estadística descriptiva para las medidas de tendencia central y de dispersión (mediana, intervalo intercuartiles, moda, etc.).

La segunda es de orden epistemológico y mucho más importante, al hablar de “IC 95% de la curva normal” de Gauss, muchos han supuesto que estamos hablando de valores “clínicamente normales”. Sin embargo, podemos darnos cuenta fácilmente de que no es así si consideramos, por ejemplo, la distribución

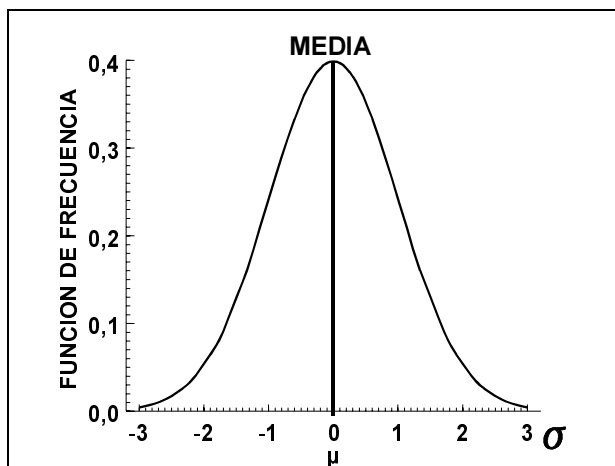


Fig. 1. Curva de Gauss estándar ( $\mu = 0$ ,  $\sigma^2 = 1$ ).

de los valores plasmáticos de LDL-colesterol. Muchos de los pacientes que estarían dentro del IC95 tienen un riesgo aumentado de eventos cardiovasculares o muerte, y como ya discutimos en otras cartas inicialmente se consideraba normal un umbral  $< 130$  mg dl, luego  $< 100$  mg dl y actualmente sospechamos que no existe ningún umbral definido en nuestra sociedad actual y por lo tanto cuanto más bajo, menor es el riesgo cardiovascular.

Esto sucede de esta manera en la práctica médica, debido a que las mediciones estadísticas de las variables biológicas no pueden por sí y ante sí reconocer el "juicio que significa estar enfermo", sin que ese "valor" haya surgido en algún momento en la conciencia del hombre. Es decir, el "gold standard" de las variables de laboratorio presentes y futuras es la conciencia del hombre que se experimentó distinto de su estado normal; es por ello que *no son las mediciones de las variables biológicas las que determinan por sí que la persona está enferma, sino que es la noción de enfermedad del paciente la que asigna valor de anormalidad a una medición biológica.*

### El "hombre tipo" de Quetelet

Tratemos de rastrear cómo surgió esta confusión. Cuando nos acercamos a mediados del siglo XIX aparece una figura, la del astrónomo real de Bélgica Adolfo Quetelet, como el impulsor principal del movimiento estadístico.

Hasta ese momento las estadísticas sólo se usaban para mostrar "estabilidades", como la cantidad de juicios, las diferentes clases de crímenes, la cuantificación de las enfermedades; en suma, estaban hechas por funcionarios estatales que llevaban registros minuciosos. Quetelet pensó en la utilización de la estadística no como un funcionario sino como un hombre de ciencia, como un astrónomo, que buscaría en las características físicas y de la conducta de sus conciudadanos regularidades dignas de las regularidades de los astros que medía en su observatorio.

Es casi seguro que Quetelet participaba de la concepción determinista del mundo de la mayoría de los científicos contemporáneos, y esa concepción determinista estaba siendo amenazada por la nueva ciencia estadística, ya que la determinación de la relación directa de "causa y efecto" era desafiada por la reciente teoría de la probabilidad.

Era necesario encontrar "leyes sociológicas" que fueran semejantes a las "leyes naturales" de los astrónomos. Quetelet estaba enamorado de los números y le gustaba sacar conclusiones; veremos cómo sus conclusiones siguen influyéndonos en el siglo XXI.

Quetelet conocía muy bien la "ley de errores" desarrollada a principios del siglo XIX, porque la aplicaba en su campo de estudio que eran las mediciones astronómicas. La representación gráfica de dicha ley es la ya discutida "curva acampanada" de la llamada "distribución normal" o "distribución de Gauss" alrededor de la media.

Había dos caminos para llegar a esa curva. El más antiguo, de comienzos del siglo XVIII, se establecía como límite de una tirada de monedas o como una distribución de binomios (la proporción que sale cara de una muestra  $n$  de monedas lanzadas al aire, repetidas muchas veces). El otro camino era la repetición de medidas de posición de un astro que realizaban los astrónomos, que presentaba una desviación de los errores de observación, establecida matemáticamente por Gauss.

Ahora bien, la curva está caracterizada por dos variables: la media y algún grado de dispersión. Si todas las medidas se agrupaban cerca de la media, la medición era confiable, y según aumentaba su grado de dispersión se volvía menos confiable. Pero aun en esta última situación, la medida real –porque no debemos olvidar que estamos estimando medidas reales de la naturaleza– más probable era la media.

La medida de dispersión que se usó en todo el siglo XIX, ya que nuestro conocido "desvío estándar" fue acuñado por Karl Pearson recién en 1894, fue el "error probable" del también astrónomo F. W. Bessel en 1815. La idea central consiste en que el "error probable" divide las mediciones en dos clases igualmente probables: la mitad de las mediciones errará por exceso y la otra mitad por defecto.

Quetelet, haciendo un juego de prestidigitación, aplicó la misma curva de distribución gaussiana de errores a fenómenos biológicos y sociales en los que el término medio (la media o el promedio actual) no va a ser nunca más una cantidad real. ***O sea con un pase de manos transformó la media que era una cantidad real (la medida real más probable), en un concepto abstracto estadístico.***

Por qué decimos que Quetelet hizo de la media (promedio) un concepto abstracto; pensemos en un hombre promedio ("hombre tipo" como diría Quetelet) que pudo haberse casado 1,22 veces o tenido 2,7 hijos, para darnos cuenta rápidamente de lo absurdo que resulta considerarlos cantidades reales; el hombre promedio es sólo una cómoda ficción.

Sin embargo, para Quetelet el "hombre medio" o el "hombre tipo" hablaba de las características de un pueblo o de una nación como tipo racial. Si antes se definía a una nación por su lugar geográfico, su historia, su lengua o su cultura; ahora habría que agregarles las características físicas y morales que compendiarían a esa nación. (3)

En Bruselas, en 1844 escribió: *"Otra cuestión de suma importancia se presenta aquí. Puede uno preguntarse si en un pueblo existe un 'homme type', un hombre que represente a ese pueblo por su estatura y en relación con el cual deben considerarse todos los demás hombres de la misma nación como desviaciones más o menos grandes. Las cifras que uno obtenga al medir a estos últimos se agruparán alrededor de un término medio, de la misma manera que los números que uno obtendría si ese mismo hombre tipo hubiera sido medido nume-*

### **rosas veces mediante métodos más o menos imprecisos”.**

En este ensayo que está dirigido a sus alumnos más distinguidos para celebrar los 30 años de profesorado, transmuta claramente la curva gaussiana de distribución de errores de medición de cantidades físicas desconocidas, a la curva de dispersión de los atributos de una media abstracta e ideal de una población. Utilizando un pensamiento por analogía y aprovechando que estas propiedades podían someterse a las mismas técnicas formales algebraicas, las convierte en propiedades reales.

Utilizando como fuente una publicación en el *Edinburgh Medical Journal* sobre la medida de la altura y el tórax de más de 5.000 soldados escoceses, llegó a la conclusión de que el “tipo medio” de tórax era de 92 cm de perímetro, y que esto era como medir un solo escocés de 92 cm de tórax con un “error probable” de 33,34 milímetros.

Según lo expresó al año siguiente en una obra popular, si *“una persona poco práctica en medir el cuerpo humano midiera repetidamente a un soldado típico, 5.738 mediciones hechas en un individuo no se agruparán ciertamente con más regularidad... que las 5.738 mediciones hechas en los soldados escoceses; y si se nos presentaran las dos series sin ninguna especificación particular, nos veríamos muy confusos para establecer qué serie se tomó de los 5.378 soldados diferentes y cuál se obtuvo de un solo individuo aplicando menos habilidad y medios más toscos de medición”*. (3)

Esta descripción ficticia de Quetelet podemos llamarla hipótesis del “sastre toscó” o quizá de “la cinta métrica aleatoriamente variable”.

Ahora, ¿cómo legitima Quetelet el paso de la abstracción aritmética de un promedio ficticio a un número objetivo y real que describe a la población?

Él da a entender que si los errores de una medición de un solo hombre concreto tienen una distribución gaussiana, y si ahora utilizando la misma técnica y conociendo que estamos midiendo a diferentes hombres, también nos da una curva de distribución normal de Gauss satisfactoria, entonces podríamos inferir, como en un silogismo, que el “tipo medio” –la media o el promedio actual–, es por lo tanto un valor verdadero y existente, una propiedad, ya no de una persona, sino de una entidad colectiva llamada población.

Pero, ¿se ajustaba el fenómeno de medición de individuos distintos a la curva de distribución normal como postulaba Quetelet?

Él mismo aplicó inmediatamente su distribución a la altura de los reclutas franceses, y cuando los resultados refutaron su teoría, encontró una explicación

“ad hoc”, argumentando que los reclutas franceses hacían fraude fingiendo baja estatura... porque en ese país existía el servicio militar obligatorio (?).

Es difícil conocer hasta qué punto los datos recogidos por Quetelet respondían realmente a la “curva de distribución de errores” de Gauss, ya que en esa época no existían las pruebas para conocer si una distribución era verdaderamente “gaussiana”.

Durante muchos años a nadie se le ocurrió poner a prueba si los fenómenos de las poblaciones se ajustaban a la curva de distribución de errores de Gauss, simplemente porque nadie se planteaba esta cuestión en esa época. Los autores alemanes escépticos, como Lexis, propusieron las primeras pruebas 30 años después.

Quetelet no se quedó aquí, ya que siguió trabajando en las mediciones estadísticas del cuerpo humano, creó la *“antropometría”* e impulsó el desarrollo de los métodos estadísticos en la medicina.

Aunque su nombre nos suene raro a los médicos asistenciales, todos implícitamente lo homenajeamos cuando diariamente utilizamos el “Índice de Quetelet”, que no es otro que nuestro muy conocido “Índice de Masa Corporal”.

### **Conclusión**

Ahora sí sabemos definitivamente que no existe una única dispersión de las distintas medidas biológicas en las poblaciones, y que las distribuciones deben ser reconocidas como se presentan empíricamente en la naturaleza.

Alguien ha comentado que la utilización de la “curva de errores” de Gauss para la descripción de la distribución de las variables biológicas en las poblaciones más que a una “curva de errores” se parece a una “comedia de errores”, ya que los estadísticos-matemáticos pensaban que era un hecho empírico comprobado por las observaciones de los biólogos, y éstos por el contrario pensaban que era una ley matemática enunciada por aquéllos.

Es ya tiempo de que terminemos con esta confusión.

**Hernán C. Doval**

### **BIBLIOGRAFÍA**

1. Sólo como curiosidad escribimos la ecuación desarrollada por Gauss:  $Y = [1 / \sqrt{(2\pi\sigma^2)}] \cdot \exp[-0,5 \cdot (x - \mu)^2 / \sigma^2]$ , donde exp indica antilogaritmo natural.
2. Tessler, J. Módulo: Estadística descriptiva de datos cualitativos, del curso virtual “La investigación clínica en la práctica médica”. GEDIC.com.ar.
3. Hacking, I. La domesticación del azar. La erosión del determinismo y el nacimiento de las ciencias del caos. Gedisa Barcelona; 1995.