

# LA ESTADÍSTICA Y EL CASTELLANO. LA TERMINOLOGÍA ESTADÍSTICA EN LA NUEVA EDICIÓN DEL *DICCIONARIO DE USO DEL ESPAÑOL*

**José A. Tapia Granados**

New School University, New York USA



Para quienes tienen interés en cuestiones lingüísticas y gustan de consultar diccionarios y enciclopedias, la publicación de una segunda edición del *Diccionario de uso del español* ha sido sin duda un acontecimiento. Somos muchos los que hemos usado durante años la primera edición del *Diccionario de uso del español* de María Moliner, todo un monumento a la sabiduría y a la voluntad de su autora, que llevó a cabo por sí sola una labor ingente que otros han realizado, con peores resultados, con toda clase de apoyos institucionales. Pero el tiempo no pasa en balde y como toda obra humana los diccionarios también se ponen viejos. El de María Moliner lo estaba y todos hemos recibido con agrado esta nueva edición, revisada y actualizada.

Hacer una evaluación seria y rigurosa de un diccionario como este del que la editorial Gredos publica ahora la segunda edición es una tarea difícil. Un diccionario es una obra en la que hay que aplicar sistemáticamente criterios diversos de inclusión, exclusión y definición. Hacer una evaluación acabada de un diccionario tan extenso como este exige muchas horas de ardua lectura o muchos años de consultas frecuentes. Los siete lustros transcurridos desde la primera edición del *Diccionario de uso del español* han demostrado sobradamente que María Moliner utilizó criterios que hicieron de su diccionario una obra muy útil para millones de usuarios de la lengua española. Los próximos años habrán de indicar en qué medida quienes han llevado a cabo la segunda edición del *Diccionario de uso del español* han acertado en su labor de actualización y revisión.

Lo que se comenta aquí es el tratamiento de la terminología estadística en la nueva edición del *Diccionario de uso del español*. Dicho sea de paso, ahora que la autora de la edición original es solo inspiradora de la tarea lexicográfica plasmada en el diccionario hay que preguntarse hasta qué punto es apropiado seguir llamando al diccionario «María Moliner». La nueva edición, en cuya cubierta aparece con caracteres bien visibles el nombre de la autora de la primera edición, dice recoger en lo fundamental la orientación general y gran parte del contenido, definiciones y ejemplos de aquella. Viene además avalada por el prestigio de la editorial Gredos y por la presentación elogiosa de Manuel Seco. Pero queda un tanto confuso quién o quiénes son los responsables finales de lo que dice el libro. La nota anónima que consta en la página xv dice que Joaquín Dacosta coordinó y dirigió «con eficacia y rigor» la redacción final de esta segunda edición. Aparecen también varios nombres de colaboradores en diversa tareas y en varios temas (se mencionan en concreto psicología, medicina, informática, terminología militar y otros, pero no matemática). Sin embargo, ninguno de esos nombres corresponde a los titulares del *copyright* que vemos en la página de créditos. En la primera edición la titular del *copyright* era María Moliner. Esperemos que la discrepancia entre la titularidad de los derechos de autor (probablemente debida a cuestiones comerciales y herencias) y la autoría real del diccionario no haya sido perjudicial para esta edición y no lo sea para futuras ediciones.

En una evaluación provisional que solo el tiempo podrá confirmar o desmentir, parece que estamos ante un diccionario que tiene muchas de las características que convirtieron en clásico a su antecesor. Los cambios que más saltan a la vista son la ordenación de las entradas con un criterio alfabético estricto —no por familias etimológicas como en la primera edición— y la consiguiente unificación tipográfica de las entradas, cambios que según se lee en las páginas iniciales del libro ya había recomendado María Moliner antes de morir y que facilitan considerablemente el uso del libro.

Pero entremos al tema. Se trata de evaluar el contenido del diccionario en lo que respecta a la terminología estadística. Obviamente estamos ante un diccionario general y por tanto sería absurdo buscar aquí definiciones muy precisas de términos especializados de la estadística, o términos estadísticos especializados que no tienen uso en el lenguaje general. Esto ya muestra hasta qué punto la elaboración de un diccionario implica una infinidad de decisiones de inclusión y de exclusión, muy difíciles a veces. Suele ser muy problemático juzgar si un término ha trascendido las fronteras de una jerga local para integrarse en un lenguaje profesional, técnico o sectorial o si ha pasado de este al habla habitual. Son al fin y al cabo juicios de hecho —no juicios de valor—, pero para poder decidir en cada caso haría falta información estadística de la que en general no se dispone. Por ejemplo, ¿qué proporción de los hispanohablantes ha oído hablar de la media geométrica? ¿Y cuántos entienden lo que es una muestra sesgada? Por otra parte, decidir si «estocástico», «aleatorio», «curtosis» o «ley de los grandes números» son o no expresiones «cultas» que *deben* ser incluidas en un diccionario general es más bien un juicio de valor. En este tipo de problemas la decisión queda al buen juicio del compilador.

En esta segunda edición del *Diccionario de uso del español* el campo al que corresponde un término o una acepción especializada aparece abreviado al comienzo de su definición. En la lista de abreviaturas usadas (págs. xxxviii-xlii del tomo i) hallamos abreviaturas para apicultura, artillería, astronomía, derecho, economía, equitación, escultura, esgrima, filosofía, geometría, matemáticas y óptica, por citar solo algunos campos que cuentan con abreviatura específica. En la lista no figura en cambio ninguna abreviatura correspondiente a estadística y, exceptuando la geometría, tampoco se encuentra subdivisión alguna (por ejemplo álgebra, topología o teoría de conjuntos) del amplio campo de la matemática, del que la estadística puede considerarse un subconjunto. Esta «penuria matemática» contrasta con la abundancia en la lista de abreviaturas de disciplinas biológicas y médicas (anatomía, biología, bioquímica, botánica, cirugía, fisiología, farmacia, medicina, veterinaria y zoología) e ingenierías y artes técnicas (agricultura, aeronáutica, arquitectura, artes gráficas, artillería, cantería, carpintería, cinematografía, fortificación, fotografía, informática, etc.), así como campos que generalmente serían considerados «menores» (astrología o cetrería, por ejemplo). Que en el diccionario no se clasifiquen específicamente como tales los términos o acepciones correspondientes a la estadística puede dar la impresión de que a este campo científico no se le ha prestado especial atención al preparar esta nueva edición. Por desgracia esta impresión se confirma al consultar algunos términos estadísticos de uso corriente. El visionario H. G. Wells dijo una vez que, algún día, para ser un buen ciudadano el razonamiento estadístico sería tan importante como saber leer y escribir. La omisión en esta nueva edición del diccionario de muchos términos estadísticos que ya son de uso corriente y la falta de definiciones o las definiciones imprecisas de otros muchos términos estadísticos quizá son *muestra* —véase aquí cómo la estadística sale por todas partes en el habla corriente— de hasta qué punto estamos lejos en el mundo hispanohablante del ideal del autor de *La guerra de los mundos*.

La estadística puede considerarse una parte de la matemática aunque estrictamente no lo es, ya que la estadística tiene aspectos experimentales mientras que la matemática es una disciplina meramente lógico-deductiva. En cualquier caso la estadística, que se ocupa del estudio abstracto (cuantitativo) de las grandes colecciones de datos, está directamente relacionada con la parte de la matemática denominada teoría de la probabilidad o teoría probabilística. Y esta es una puerta de entrada al tema, porque, por ejemplo, no hallamos en el diccionario una definición matemática de *probabilidad*. Por supuesto que dar esa definición es tarea muy difícil en la que las escuelas frecuentista y bayesiana llevan decenios debatiendo, pero un diccionario general como este no necesitaría entrar a tales sutilezas. Por cierto que el adjetivo *bayesiano* (del matemático Thomas Bayes, 1702-1761) no aparece en el diccionario (donde sí encontramos en cambio epónimos como *estajanovista* —de A. G. Stajanov—, *maltusiano* —de Thomas Malthus— o *keynesiano* —de J. M. Keynes—). Pero volviendo a *probabilidad*, sería conveniente que el diccionario dijera, por ejemplo, que en matemática o estadística se entiende por probabilidad un número entre cero y uno que expresa la frecuencia con la que ocurre un fenómeno en una serie de ocasiones en las que el fenómeno puede o no ocurrir; o la creencia subjetiva en que un fenómeno va a ocurrir en una ocasión dada, siendo el valor cero correspondiente al suceso que ocurre con

frecuencia nula o del que se tiene absoluta seguridad que no ocurrirá, mientras que el valor uno indica un suceso que siempre ocurre o del que tenemos seguridad absoluta que ocurrirá. Una definición así sería mejor que la ausencia de definición matemática que encontramos.

Dos términos básicos en la teoría probabilística son *aleatorio* y *estocástico*, cuyos significados son prácticamente indistinguibles. Ambos están recogidos en esta edición del diccionario (*aleatorio* ya lo estaba en la primera) con definiciones a las que no hay nada que objetar.

En estadística suelen considerarse básicas tres tipos de medidas, razones, proporciones y tasas. El diccionario explica en la 5ª acepción de la entrada *razón* que se trata de la relación establecida por división entre dos magnitudes, expresada numéricamente. También dice que en matemáticas razón significa el cociente indicado de dos números y se expresa por escrito con raya de quebrado o con el signo «:» (2/3, 2:3), leído, «dos tercios, dos partido por tres» o «dos es a tres».

De pasada puede mencionarse que la redacción de esta definición deja que desear, ya que incluye en un paréntesis el referente que luego se desarrolla fuera de él (más lógico hubiera sido que el paréntesis abierto delante de 2/3 se cerrara después de «dos es a tres»). El uso estadístico del término *razón* es habitual, por ejemplo en expresiones como «la razón profesores-alumnos en el conservatorio es 6:60», que simplemente indica que hay 10 alumnos por cada profesor ( $6:60 = 1:10$ ); o «la razón capital-producto es 0,02», que indica que por cada unidad de capital se obtienen 50 unidades de producto ( $1:50 = 0,02$ ). En ciencias sociales este uso estadístico de *razón* es muy frecuente. Sin embargo, los economistas hispanohablantes, sometidos a una fuerte influencia del inglés y casi siempre muy poco interesados en asuntos lingüísticos, a menudo usan una jerga en la que saltan cada dos por tres las palabras inglesas. Así en textos económicos, redactados a menudo en *español*, es cada vez más frecuente encontrar la palabra *ratio* en vez del término castellano *razón*. Y, por desgracia, en esta edición del *Diccionario de uso del español* encontramos una entrada para el término *ratio*, referido a economía y definido como «índice numérico que establece una proporción entre dos elementos contables, por ejemplo entre activo y pasivo». Esta definición, que tampoco es especialmente precisa, no es más que uno de los usos estadísticos de *ratio* en inglés. Ciertamente, como dice en esa entrada, *ratio* viene del latín «ratio, -onis», cálculo, pero en este caso pasando por el inglés, ya que *ratio* originó en castellano *razón* de la misma manera que *mutatio* dio *mutación*, *populatio* dio *población*, *titio* dio *tizón* y así sucesivamente. Dicho de otra manera, *ratio* es un anglicismo a mi juicio innecesario, que no debería haberse incluido en el diccionario a menos que se señalara su dudoso pedigrí y la conveniencia de sustituirlo por *razón*.

Del término *proporción* el diccionario dice que se refiere a la relación entre las dimensiones de una cosa o las de dos o más cosas y añade que en matemáticas se refiere a la igualdad de dos razones. Hubiera sido aquí conveniente añadir que en sentido estadístico, si una razón es cualquier expresión de la forma  $a/b$ , dicha razón es también una proporción siempre que  $a$  esté contenido en  $b$ . Así si las camas del departamento pediátrico son 20 de

las 80 camas de un hospital podemos decir que la proporción de camas pediátricas en ese hospital es 20/80, o 1/4, o 0,25, o 25%. En sentido matemático preciso toda proporción es una razón, aunque lo inverso no es cierto. La razón 7:2 del número de diputados varones al número de diputadas de Bolulandia no es estrictamente una proporción, ya que 7:2 no expresa la relación de la parte al todo (otra cosa sería si dijéramos que la proporción de mujeres entre los diputados de Bolulandia es 2/9, o sea, 22,2%). Por otra parte el diccionario tampoco indica que las proporciones se expresan muy a menudo en forma de porcentaje, como en el ejemplo indicado.

El término *porcentaje* viene en el diccionario con una etimología incorrecta, ya que dice que procede del inglés «percentage». Pero en inglés se escribe *percentage*. En esto la segunda edición desmejora la primera, en la que María Moliner decía que porcentaje viene de «por ciento» y no daba una etimología equivocada.

Un campo fundamental de la estadística es el de las llamadas medidas de centralización y de dispersión, que son las que se usan para resumir o describir una colección más o menos numerosa de datos numéricos. En inglés las medidas de centralización se denominan *averages* e incluyen lo que en castellano llamamos media, mediana y moda, a las que en conjunto suele denominarse «promedios» (en Cuba he oído «centroides» con este mismo significado; el término está bien construido y sería usable si tuviera más difusión). De forma que en sentido general, las medidas de centralización son lo mismo que los promedios y la mediana y la moda son «tan promedios» como la media. El tratamiento de estos términos en esta segunda edición del *Diccionario de uso del español* no es especialmente bueno. En *media* se remite inmediatamente a *media aritmética*, que se define correctamente como una subentrada matemática. También como tales se definen *media diferencial* y *media proporcional* (ambas expresiones a mi juicio no tienen uso alguno en el lenguaje corriente). Sin embargo, no se define ni la *media geométrica* (que no es más que la raíz enésima del producto de los  $n$  datos numéricos) ni la *media (h)armónica*, que es el inverso de la media aritmética de los inversos de los datos). Los usos de la media geométrica no son ni mucho menos excepcionales y hubiera sido conveniente que su definición se incluyera en el diccionario. La *media ponderada* se usa muy a menudo en diversos campos científicos, pero tampoco se define en el diccionario. *Promedio* sólo viene definido como sinónimo de media aritmética, lo que, como ya dije anteriormente, no es aceptable, ya que cualquier media o incluso la mediana o la moda pueden ser promedios. En cuanto a *mediana* (el valor que divide justamente en dos partes iguales a una distribución) y *moda* (el valor más frecuente de una distribución), ninguna de estas dos entradas se halla definida en su sentido estadístico.

Si las medidas de centralización son básicas en estadística, las medidas de dispersión no lo son menos. Entre ellas la principal es la que (por influencia del francés *écart-type*) antes solía denominarse «desviación típica» aunque ahora (por influencia del inglés *standard deviation*) cada vez más se conoce por desviación estándar, o desvío estándar en algunos países de América. Ni *desviación típica* ni *desviación estándar* constan en el diccionario y en *desvío* tampoco se hace referencia alguna a su posible acepción estadística. Tampoco

encontramos *varianza*, que es como la mayor parte de los textos científicos en castellano se refieren a la media aritmética de los cuadrados de las desviaciones de cada valor respecto de la media aritmética de los datos (por lo que a veces también se denomina *desviación cuadrática media*). La desviación estándar es la raíz cuadrada de la varianza. En algunos textos en castellano se ve *variancia* en vez de *varianza*, pero esta grafía se usa muy poco, pese a ser la recomendada por la Real Academia.

También son términos estadísticos de uso cada vez más frecuente los que se refieren a particiones de una distribución. Seguramente más de una madre sabe lo que quiere decir que su hijo, bastante delgado, está en el percentil 15 de peso para la edad. Los percentiles se obtienen dividiendo una distribución en cien partes iguales. Así la mediana corresponde al percentil 50 y si un niño está en el percentil 15 de peso para la edad quiere decir que de cada 100 niños de su edad 85 pesan más que él, y 15 menos. De la misma manera, si dividimos una distribución en cuatro partes iguales cada valor divisorio corresponde a un cuartil, si la dividimos en cinco, a un quintil o quintilo y si la dividimos en diez, a un decil o decilo. En los textos científicos en castellano hay gran diversidad de estos términos, que también se ven a veces en femenino, como cuartilas, quintilas, decilas y percentilas. También se usa a veces el término *centil* en vez de *percentil*. En cualquier caso, en esta edición del diccionario no constan ni *cuartil* ni *decil* ni *percentil* ni *centil* ni ninguna de las variantes mencionadas. *Quintil* sí aparece, pero no en su acepción estadística.

El término *muestreo* sí que consta en esta nueva edición del *Diccionario de uso del español* y además se define específicamente referido a su significado estadístico: «En estadística, selección de una parte representativa de un conjunto para estudiarla y sacar conclusiones que afectan a todo el conjunto». Esta definición, a la que no hay nada que objetar, no tiene complemento en el verbo correspondiente, *muestrear*, que no aparece en el diccionario, aunque se usa mucho en los textos técnicos.

La palabra *sesgo* tiene una acepción estadística bien definida que tampoco recoge el diccionario. «Como la encuesta solo se hizo en el barrio de Salamanca habrá sesgo en la proporción muestral» significa que la proporción muestral se desviará considerablemente de la proporción poblacional, siendo la población en este caso el total de los residentes en Madrid. Tampoco se explica la acepción estadística del verbo *sesgar* («la muestra está sesgada»). En *población* el diccionario explica los significados habituales e incluso remite a estadística en uno de los catálogos, pero no da el significado estadístico del término, que se refiere al conjunto del que se extrae una muestra y que por tanto puede corresponder a un conjunto de seres humanos, pero también a granos de cereal o piezas salidas de una cadena de producción. El adjetivo *poblacional*, de uso frecuentísimo en estadística, ni siquiera consta en el diccionario. En sentido estadístico *población* es sinónimo de *universo*, término del que tampoco se define su significado estadístico. La contraposición entre lo muestral y lo poblacional es el fundamento de la inferencia estadística, gracias a la cual se llega a conclusiones sobre la totalidad (la población o universo) a partir del conocimiento parcial (la muestra). Así el *valor poblacional* o *parámetro* se estima a partir del *valor muestral*, también llamado *estadístico* o *estadígrafo*. Por supuesto que el diccionario no explica estos significados estadísticos de

*estadístico, valor muestral y parámetro*; el término *estadígrafo* ni siquiera aparece.

Todo lo dicho creo que justifica que en lo que respecta a la terminología estadística puedan hacerse bastantes reparos a esta nueva edición del *Diccionario de uso del español*. En muchos aspectos la estadística ya forma parte de la cultura general y los compiladores de este «María Moliner» no han contribuido ni a favorecer ni a consolidar esa presencia en el ámbito lingüístico del castellano. Quizá en sucesivas ediciones del diccionario pueda repararse este defecto.

## NOTAS

1. Ni *varianza* ni *variancia* ni *desviación estándar* se definen en esta edición del *Diccionario de uso del español*, pese a ser conceptos ampliamente usados hoy día en casi todos los campos científicos. El coeficiente de variación y la desviación media son medidas de dispersión quizá menos utilizadas y que tampoco aparecen en el diccionario. Mucho menos grave es que no encontremos *curtosis* o *kurtosis*, otro concepto estadístico referente a la dispersión unidireccional de los datos, este sí, mucho más especializado.

2. En su *Diccionario crítico de dudas inglés-español de medicina* (Madrid, McGraw-Hill, 2000) Fernando Navarro defiende el uso de *variancia* y considera que el término *varianza* es un anglicismo, con criterios a mi juicio equivocados, como he explicado en la nota titulada «¿Varianza o variancia?», publicada la *Revista Panamericana de Salud Pública* (vol. 8, No. 5, 2000, pp. 368-370).