DESVIACIÓN, DESIGUALDAD, POLARIZACIÓN: MEDIDAS DE LA DIVERSIDAD SOCIAL

Modesto Escobar
Universidad de Salamanca
Instituto Juan March de Estudios e Investigaciones

RESUMEN

En este artículo se presenta una distinción entre los conceptos de desviación, desigualdad y polarización y se propone una familia de coeficientes aplicables a variables con valores limitados, es decir, con valores mínimo y máximo, a fin de medir la desviación de las distribuciones en términos porcentuales. Entre los coeficientes propuestos, el más simple y útil es el coeficiente de variación ajustado, que implica la adición de una serie de propiedades deseables al conocido coeficiente de variación de Pearson para las variables limitadas. Finalmente, se presenta un par de aplicaciones donde se ponen de relieve las ventajas del coeficiente presentado.

El presente artículo contiene una serie de propuestas sobre la medición de la variación, uno de los aspectos más controvertidos en los estudios empíricos sobre la diversidad, entre los cuales se encuentra la medición de la desigualdad, que suscita una polémica inacabada que sigue despertando el interés de economistas y sociólogos. El tema de la desigualdad social es central en el estudio de la Sociedad: desde las concepciones jerárquicas de Platón al discurso sobre el origen de la desigualdad de Rousseau ya existe la preocupación por la diversidad en el pensamiento presociológico. Posteriormente, en los clásicos la preocupación por el estudio de las divisiones en la sociedad es obvia y puede ser sintetizada con los conceptos de clase en el caso de Marx, *status* en el de Weber y solidaridad mecánica/orgánica en el caso de la obra de Durkheim. Nisbet señala el *status* como una de las cinco ideas-clave de la tradición sociológica. Las similitudes y desigualdades dentro y entre sociedades deben ser estudiadas por los sociólogos, y los metodólogos tienen el reto de proporcionar herramientas precisas que permitan medir lo diverso en la sociedad.

Pero si importante para el estudio de la sociedad es encontrar instrumentos que nos faciliten el estudio de la diversidad, no lo es menos desde un punto de vista estrictamente metodológico: a menudo se confunde la estadística con la ciencia de los porcentajes y las medias, porque con mucha frecuencia sus estudios divulgativos sólo emplean estos estadísticos para contribuir a la sencillez y comprensión de lo que se quiere presentar. Sin embargo, tan importante para dar una correcta imagen de la sociedad es ofrecer los promedios de las variables como la forma en cómo están distribuidas. Y, de modo similar, la estadística avanzada está construida básicamente a partir del concepto de varianza. En general, la mayor parte de las técnicas que sirven para la explicación causal entre variables utilizan la expresión «porcentaje de varianza explicado».

Antes de desarrollar una serie de estadísticos cuyo fin primordial es cuantificar la diversidad presente en poblaciones estudiadas a través de muestras, es conveniente precisar unas notas conceptuales en el estudio de la diversidad. Para ello es importante distinguir tres conceptos que, aunque similares para un profano, son tres aproximaciones distintas de la diversidad. Estas tres dimensiones de la diversidad son: la desviación, la desigualdad y la polarización.

Por *desviación* debe entenderse lo que los sujetos se alejan de un determinado punto de referencia, el más común del cual es la media aritmética. Quiere ello decir que existe un patrón con el que se comparan todos los sujetos y la medida de desviación será tanto mayor cuanto los individuos estén más alejados de la referencia elegida. En el concepto de *desigualdad* no hay un único punto de comparación, sino que cada sujeto ha de compararse con el resto de individuos de su población. Por último estaría el concepto de *polarización* (o su contrapuesto el de concentración), que indica hasta qué punto los valores de las variables están cerca o lejos no de una sola referencia, ni de todos los valores, sino de dos puntos que son los extremos de la distribución.

Es evidente que estas tres concepciones de la diversidad están estrechamente relacionadas entre sí, pues los distintos estadísticos que existen para medirlas están condicionados, de modo que un aumento de una medida de la desviación implica un aumento en la medida de la desigualdad, por lo que no es infrecuente utilizar estadísticos de una índole para indicar lo que otro tipo de estadístico expresa mejor.

En este artículo se propone una familia de estadísticos que encaran el tema de la diversidad social desde el punto de vista de la polarización, abarcando no sólo a medidas de desviación, sino también a medidas de desigualdad.

Una de las medidas más utilizadas en análisis complejos de la estadística es la *varianza,* que es el promedio de las desviaciones al cuadrado de los valores de una variable con relación a la media. Para ver la especificidad de este estadístico, imagínese un reparto de cuatro pesetas indivisibles (ya no existen ni céntimos ni reales) y supóngase que hay cuatro personas a las que repartir la suma indicada. Además de la situación igualitaria de que cada cual obtuviera una única peseta, podrían ocurrir cuatro fórmulas de desigualdad monetaria: *a)* que una sola persona se quede sin nada, por lo que otra se quedaría con dos pesetas; que dos personas se queden *sin blanca,* en cuyo caso cabrían dos subsituaciones: *b)* que las otras dos personas se repartan a medias las cuatro pesetas, y *c)* que la segunda cogiera una peseta y la primera se apropiara, además de la suya, de las dos pesetas que no pasaron a propiedad de las dos últimas personas; y *d)* que tres personas queden sin nada y la primera persona se quede con las cuatro piezas.

La situación igualitaria (todos se reparten una peseta) arroja una varianza nula: todo el mundo se mantiene en el punto medio; no hay variación. En la situación *a)* hay dos personas que se desvían una unidad de la media. Como en conjunto hay cuatro personas en nuestro análisis, el promedio de la desviación es media peseta. En *b)* la varianza toma el valor unidad porque los cuatros sujetos se desvían de la media una unidad. Las dos últimas situaciones son peculiares de la varianza: En la penúltima, un sujeto obtiene dos pesetas por encima de la media; por ello, al promediar su desviación, en lugar de dos, cuenta como cuatro y de esta forma el conjunto de la varianza es uno con cinco (seis entre cuatro personas). En la última, el sujeto que se queda con las cuatro monedas contribuye al sumatorio de la varianza con tres al cuadrado (nueve) y de esta forma el resultado del cálculo de este estadístico es 3, es decir, el doble del anterior.

TABLA 1

Distribución de cuatro unidades

		Desigualdad			
	Igualdad	*a*	*b*	*c*	*d*
Persona W	1	2	2	3	4
Persona X	1	1	2	1	0
Persona Y	1	1	0	0	0
Persona Z	1	0	0	0	0
Media	1	1	1	1	1
Varianza	0	0,5	1	1,5	3
Desviación típica	0	0,7	1,0	1,2	1,7
Desviación media	0	0,5	1	1	1,5
Coeficiente Variación	0	0,7	1,0	1,2	1,7

Dos características saltan a la vista a partir de este ejemplo: en primer lugar, que la varianza aumenta considerablemente cuando en una variable se encuentran valores muy alejados de la media (el caso de aquel que se quedó con las cuatro pesetas) y, en segundo lugar, que las unidades en las que está expresada esta magnitud son unidades cuadradas de los valores originales de la variable. En este caso, enunciar que existe una varianza de 3 quiere decir que, por término medio, los habitantes de esta población supuesta se desvían de la media 3 pesetas al cuadrado. Por esta razón, resulta de interés reducir nuestro estadístico a unidades reales, no cuadráticas, para lo que se emplea la *desviación típica,* que se obtiene con la mera raíz cuadrada de la varianza. De esta forma, las 3 pesetas al cuadrado quedan reducidas a 1,7 pesetas mondas, redondas —no cuadradas— y lirondas.

En este contexto resulta útil recordar brevemente y sin demostraciones las propiedades matemáticas de la varianza:

1) La varianza es siempre un número mayor o igual que 0.
2) La varianza puede obtenerse mediante la sustracción de los cuadrados de la media cuadrática y la media aritmética.
3) La varianza de una constante es nula.
4) Si a una variable se le añade una constante, la nueva variable tendrá la misma varianza.
5) Al multiplicar una variable por una constante, la nueva variable tendrá una varianza C^2 veces superior.
6) La varianza de una suma de dos variables es igual a la suma de las varianzas respectivas más dos veces su covarianza.

Desventajas de la varianza —que no quedan resueltas con la desviación típica— son su carácter de magnitud absoluta y que esté medida en unidades cuadráticas de los valores sobre los que se calcula. Este último inconveniente tiene solución fácil con la obtención de la desviación típica, pues ésta proporciona la devolución de las unidades a los valores que la varianza tenía inicialmente. No obstante, siguen existiendo dos problemas: primero, resulta difícil precisar si la variable está bien o mal distribuida o, dicho de otra manera, se carece de elementos de comparación para evaluar el grado en el que se dispersan las variables en estudio —¿es mucho que la desviación típica de la variable ingreso sea igual a 3?—; segundo, resulta imposible poder comparar la dispersión entre dos variables medidas en magnitudes distintas —si, además de estudiar los ingresos, se estudiara la comida y ésta tuviese una desviación de 10, ¿se podría afirmar que esta segunda variable está más dispersa que la primera?

Este problema se solventa con las medidas relativas —o normalizadas, según Weisberg (1986)— de variación. Son medidas relativas aquellas que carecen de unidades de medición. Entre ellas, las más conocidas son, sin duda, los porcentajes y las proporciones. Ambas se obtienen con un cociente de cantidades similares y, por ello, quedan desprovistas de unidades. Si se dividen personas entre personas, el resultado no son personas, sino una proporción comparable a aquella resultante de dividir pesetas entre pesetas; en cambio, inicialmente no son comparables personas con pesetas. Recuérdese que al realizar un cociente entre magnitudes distintas debemos mencionar ambas; en cambio, si dividimos magnitudes iguales, el resultado es una medida relativa. Ejemplos claros son la velocidad: al dividir kilómetros recorridos en un viaje por las horas que se tarda, se obtiene una medida que son kilómetros/hora; al dividir alumnos en una facultad por profesores, se obtiene alumnos/profesor; pero si dividimos votantes al PSOE entre el total de votantes, la cantidad obtenida carece de magnitud.

En casi todos los manuales de Estadística aparece una medida relativa de varianza llamada *coeficiente de variación de Pearson,* que se obtiene mediante el cociente entre la desviación tipica y la media. Véase un ejemplo de su aplicación. Ante la pregunta de qué presenta mayor variación en una población: la edad o el número de hijos, el problema es que ambas variables están medidas con distintas escalas. Las unidades de la primera son años; las de la segunda, personas. Tanto las medias como las varianzas—y claramente también las desviaciones típicas— serían incomparables. Pero si se calculan los coeficientes de variación las magnitudes resultantes podrían compararse. Véase un ejemplo de una ciudad española, Madrid. La edad media de los habitantes de este municipio es de 38,2; la varianza es de 487,6 años al cuadrado; la desviación tipica es 22,1 —podría decirse algo así como que la desviación promedio por habitante en Madrid es de 22 años—. En relación con el número de hijos: las madrileñas de más de 15 años tienen, por término medio, 1,5 hijos. La desviación tipica es de 1,7 hijos. En comparación con la edad parece bastante menor; sin embargo, al utilizar el coeficiente de variación, se puede observar lo que es lógico: la edad presenta una dispersión menor que el número de hijos. El coeficiente de variación de la primera es 58 por 100, mientras el del segundo es 116 por 100. Ambos números revelan la mayor dispersión de la variable número de hijos; pero el resultado de la segunda —116 por 100— da qué pensar sobre los límites de este coeficiente de variación. En Salamanca se observa que la población es algo más joven y que las mujeres tienen más hijos por término medio. También pueden compararse las medidas de dispersión absoluta y relativa entre las dos ciudades. Se detecta que en la edad tanto la desviación típica como el coeficiente de variación son menores en la ciudad de Madrid —habrá menos niños y menos viejos—. Sin embargo, en el número de hijos la medida absoluta da mayor dispersión a Salamanca y la medida relativa a Madrid. Es preferible esta última.

Véanse ahora algunas de las propiedades de este coeficiente de variación, incluyendo también aquellas que se consideran negativas en este contexto:

1) Siempre que una variable sea constante, el valor que adopta es 0.
2) Puede adoptar valores negativos, en el caso de que la media sea negativa.

TABLA 2

Edad e hijos en dos municipios españoles

	Municipio									
	Madrid					*Salamanca*				
	N	*Media*	*Desv. t.*	*CV (%)*	*CVa (%)*	*N*	*Media*	*Desv. t.*	*CV (%)*	*CVa (%)*
Edad	3.113.818	38,2	22,1	58	46	162.737	37,2	22,6	61	47
Hijos	1.385.577	1,5	1,7	116	56	70.917	1,7	1,9	114	59

3) Cuando los valores de la media son próximos a 0, el coeficiente de variación tiene valores extraordinariamente altos, dándose el caso de ser infinito, en la circunstancia en que la media sea igual a 0.

4) Al sumar a la variable una constante C, el coeficiente de variación sufre una disminución.

5) Al multiplicar la variable por una constante M, el coeficiente de variación no se ve alterado.

6) El valor máximo que puede adquirir el coeficiente de variación en una variable positiva (es decir, sin valores negativos) es $\sqrt{n-1}$.

Por estas características, es conveniente no utilizar el coeficiente de variación cuando la variable tiene valores negativos o cuando la media es próxima a cero. En realidad, este coeficiente tiene como supuesto de uso el que la variable de trabajo se encuentre semiacotada en el límite inferior (con el valor 0), en cuyo caso la situación de mayor variación sería aquella en la que todos los sujetos menos uno tienen el valor 0 y el sujeto restante tiene el valor $n \times \bar{x}$. Volviendo al ejemplo del reparto monetario, en la situación de que sólo una persona obtenga las cuatro pesetas y las otras tres se queden sin nada, la desviación típica es $\sqrt{3}$ y el coeficiente de variación también es igual a $\sqrt{3}$, que traducido a términos porcentuales arroja un valor del 173 por 100.

Una primera medida que se puede adoptar para evitar que este cociente deje de tener el inconveniente de presentar unas cotas distintas según la distribución sería la de dividirlo por $\sqrt{n-1}$, que es su valor máximo. De esta manera, el valor anterior del 173 por 100 de variación se convertiría en el cien por cien, que indicaría que esta distribución tendría la mayor varianza posible. Así, pues, *el coeficiente de variación ajustada* presentaría la siguiente fórmula:

$$CV_{aj} = \frac{s}{x\sqrt{n-1}} \qquad [1]$$

Aun con todo, adolece esta medida de un supuesto muy débil, sobre todo cuando el *n* es alto y se supone que el valor que puede adoptar una variable es ilimitado. Un claro ejemplo sería una encuesta con 1.000 sujetos, edad media de 35 años y una desviación típica de 20 años: para el cálculo del coeficiente de variación máximo estaríamos suponiendo que habría 999 sujetos de 0 años y un sujeto con 999 × 35 años, es decir, 3.465. Ni el mítico y bíblico Matusalén cumpliría tantos años.

Otra medida de variación relativa sería resultado de la división de la desviación típica entre la máxima posible con la condición de que la variable esté acotada por ambos lados, o, dicho de otra forma, en el supuesto de que se conociera el rango de la distribución, en cuyo caso la situación más desfavorable sería aquella en la que el 50 por 100 de los casos tuviera el valor mínimo de la distribución y el otro 50 por 100 el valor máximo. Si, además, nadie pudiera tomar más de dos pesetas, el rango sería 2 y la situación de varianza máxima aquella en la que (si hay cuatro personas) dos personas poseen dos pesetas y otras dos personas se quedan sin ningún ingreso (Weisberg, 1986; 53). La media en estos casos sería:

$$\overline{X} = \frac{X_{max} - X_{min}}{2} \qquad [2]$$

Y la varianza correspondiente a la máxima polarización se ajustaría a la siguiente expresión:

$$Var_{max} = \frac{(X_{max} - X_{min})^2}{4} \qquad [3]$$

Sin embargo, este cálculo de la desviación máxima adolece de un defecto: siempre supone la media constante, aun cuando en la distribución se presente una media distinta del punto equidistante entre los valores máximos y mínimos.

Por ello, también la medida recién mencionada tiende a sobrevalorar la varianza máxima de las distribuciones. La propuesta de este artículo es la de una varianza máxima que calculándose sobre valores acotados tenga en cuenta como dato la media de la muestra, es decir, se aboga por un coeficiente que reúna las dos características deseables de las anteriores medidas.

Ante una variable acotada[1] con más de dos valores, la situación de mayor variación correspondería a aquella en la que sólo hubiera precisamente dos valores coincidiendo con los extremos.

Sabiendo que esta variable tiene una media $\overline{X}$, pueden deducirse fácilmente las proporciones de los casos de los valores mínimo y máximo:

---

[1] Se define en este contexto variable acotada como aquella variable con valores cuantitativos que tenga límites inferior (X_{min}) y superior (X_{max}) definidos.

$$\left.\begin{array}{r} p_{\min} + p_{\max} = 1 \\ \overline{X} = p_{\min} X_{\min} + p_{\max} X_{\max} \end{array}\right\} \qquad [4]$$

Y con las proporciones y los valores conocidos ya puede calcularse la varianza máxima:

$$Var_{\max} = (X_{\max} - \overline{X})^2 \, p_{\max} + (X_{\min} - \overline{X})^2 \, p_{\min} \qquad [5]$$

Fórmula que, empleando el álgebra básica, puede simplificarse del siguiente modo:

$$Var_{\max} = (X_{\max} - \overline{X})\,(\overline{X} - X_{\min}) \qquad [6]$$

Tras este desarrollo matemático, valga a continuación un ejemplo simple que exponga con variables sociológicas el significado aplicado de esta medida. Saquemos a colación la identificación ideológica medida en una escala del 1 al 7. En este caso, si la media de la muestra fuera 4, la situación de mayor polarización (varianza máxima) sería aquella en la cual la mitad de los entrevistados contestaran con un 1 (extrema izquierda) y la otra mitad con un 7 (extrema derecha).

TABLA 3

Comparación de las varianzas máximas

Media=4	x_i	p_i	$x_i p_i$	$(x_i-\overline{x})^2 p_i$	*Media=3*	x_i	p_i	$x_i p_i$	$(x_i-\overline{x})^2 p_i$
Ext. izquierda	1	0,5	0,5	4,5	Ext. izquierda	1	0,7	0,7	2,7
Ext. derecha	7	0,5	3,5	4,5	Ext. ferecha	7	0,3	2,3	5,3
			$\overline{x}$ = 4,0	s^2 = 9,0				$\overline{x}$ = 3,0	s^2 = 8,0
Var_{max} (W)	9,0				Var_{max} (W)	9,0			
Var_{max}	9,0				Var_{max}	8,0			

De forma gráfica, se apreciaría viendo que todos se desvían en términos absolutos tres puntos de la media, que elevados al cuadrado dan el resultado de 9. Ahora bien, si la media fuese de 3, en lugar de 4, dos tercios de la población estarían en el lado 1 de la ideología y sólo un tercio en el lado 7 de la escala. En el primer caso la varianza máxima no es 9, sino (7 – 3) (3 – 1) = 8, esto es, se reduce la posible variación de la distribución.

Conocida la varianza máxima, es fácil obtener también la máxima desviación típica, obteniendo la raíz cuadrada de la primera. A partir de aquí puede calcularse un par de medidas interrelacionadas: una proporción de varianza acotada (PV_a), que sea el cociente entre la varianza empírica y la máxima condicionada a la media empírica, y un coeficiente de variación acotada (CV_a), que sea la razón entre las respectivas desviaciones típicas. Sus fórmulas simbólicas serían las siguientes:

$$PV_a = \frac{s^2}{(\bar{X} - X_{min})\,(X_{max} - \bar{X})} \qquad [7]$$

$$CV_a = \frac{s}{\sqrt{(\bar{X} - X_{min})\,(X_{max} - \bar{X})}} = \sqrt{PV_a}$$

Ambas medidas presentan la deseable característica de poseer un rango entre 0 y 1. El valor mínimo se encuentra en 0 cuando la variable tiene una varianza de 0 y el valor máximo en 1 cuando la variable sólo presenta dos valores y éstos coinciden con los extremos de la distribución. Además, pueden obtenerse la una a partir de la otra, como fácilmente puede comprobarse en la última fórmula.

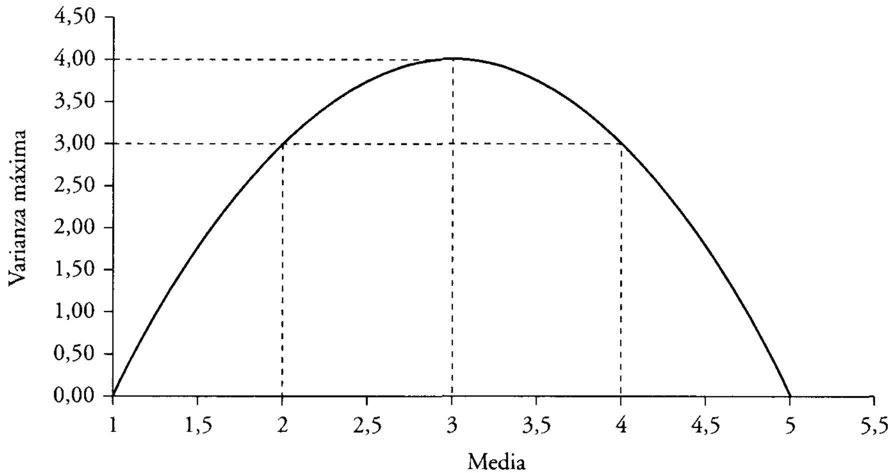
Una interesante aplicación de estas medidas, especialmente la segunda, por presentar generalmente valores mayores, es la comparación de la variabilidad entre medidas con escalas distintas, al igual que ocurría con el coeficiente de variación. Retomando el ejemplo de la edad y los hijos en Madrid. Con el CV se comparaba una variación del 116 por 100 en el caso de los hijos, frente a una del 58 por 100 para la edad. Si se utiliza el coeficiente de variación acotada[2] los porcentajes respectivos serían del 46 y 56 por 100. Es evidente que estas dos últimas cantidades responden mejor a la idea común de porcentaje.

Además, poseen como interesante característica el hecho de que una transformación lineal de la variable no altera este coeficiente, propiedad que no cumplen el resto de medidas de dispersión aplicables a variables numéricas. Quiere ello decir que si se suma o se multiplica a una determinada variable por una constante, el coeficiente obtenido con la variable transformada debe ser idéntico al calculado con la variable inicial. Ello no ocurre con la desviación típica (o varianza), pues se ve alterada al *multiplicar* la distribución por una constante, ni con el coeficiente de variación, que sufre modificación al *añadirle* una determinada constante. Esta propiedad es especialmente interesante en el campo de las ciencias sociales, donde la inmensa mayoría de las escalas son arbitrarias —¿por qué medimos la ideología de 1 a 10 y no de 0 a 20, por ejemplo?—, por lo que necesitamos medidas que resistan las decisiones sobre los valores de la escala. El caso más claro sería el de las escalas de Likert —normalmente medidas de 1 a 5—, en las que es frecuente invertir los *items* utilizando la fórmula $x' = 6 - x$. Si se utiliza el coeficiente de variación, el resultado es distinto según la utilicemos invertida o no.

[2] Como límites inferiores en ambas variables se utiliza el 0; como límite superior el 8 en el caso del número de hijos, el 98 en el caso de la edad.

GRÁFICO 1

Evolución de la varianza máxima



La variabilidad de este índice puede mostrarse de modo simple (gráfico 1) en el contexto de las escalas de Likert de 5 puntos, medidas que no son difíciles de hallar en los buenos cuestionarios. El denominador de la proporción de varianza, es decir, la varianza máxima, presenta una función parabólica invertida en relación con la media de la variable. Es lógico que si en una escala de 1 a 5 el valor medio es 1, la varianza máxima va a ser 0, porque al no haber valores inferiores a ese número sólo puede haber unos en la distribución y, por tanto, la varianza máxima posible será 0. De la misma forma, si el valor medio de la distribución es el 5, por no poder haber valores superiores, su varianza máxima tendrá valor 0. La varianza máxima de la distribución puede alcanzarse con la media de 3, pues en ese caso el 50 por 100 de los sujetos tendrán el valor mínimo (1) y el otro 50 por 100 el valor máximo (5), lo que da como resultado una varianza de 4. Con medias intermedias, la varianza máxima no puede ser ni inferior a 0 ni superior a 4. Caso de que haya en una muestra una media de 2, en cuya situación la varianza máxima se alcanzaría si sólo hubiera un 75 por 100 de 1 y un 25 por 100 de 5 (es la única combinación de 1 y 5 que da una media de 2), la varianza máxima sería de 3. En simétrica posición nos encontraríamos si la media fuese de 4; también en este caso la varianza máxima tendría el valor 3.

Los gráficos 2 y 3 nos enseñan unas curiosas propiedades del coeficiente de variación acotada frente al clásico coeficiente de variación de Pearson en una escala de cinco puntos. Es evidente que con medias similares, los coeficientes muestran unas pautas similares; pero suele ser un poco más alto el acotado. Sin embargo, la principal diferencia estriba en cuanto hay variación de medias. En

GRÁFICO 2

Evolución de los CV en función de la varianza (Media = 2)

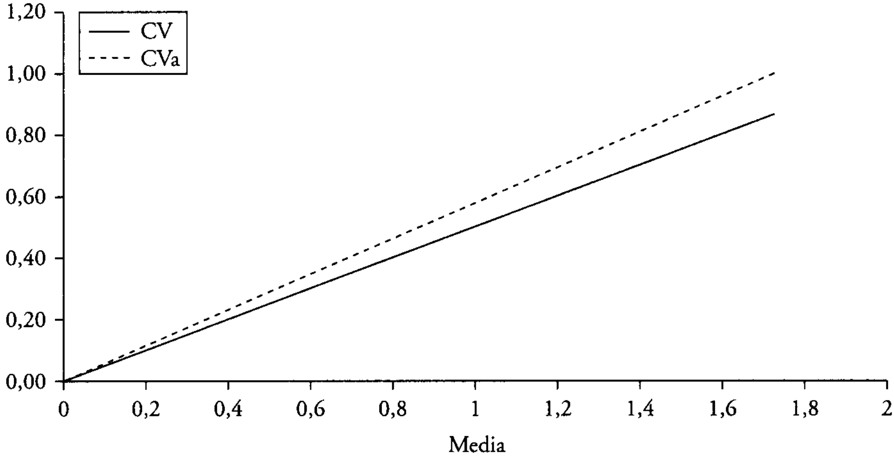
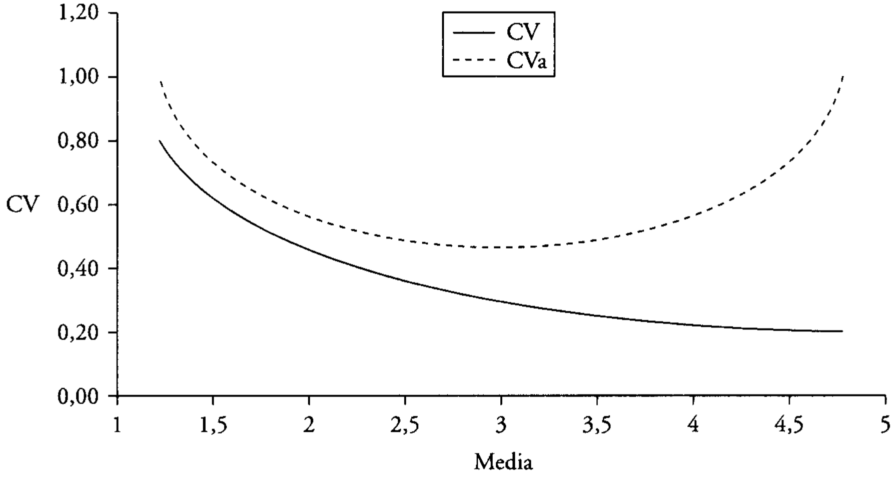


GRÁFICO 3

Evolución de los CV en función de la media (Varianza = 1)



el gráfico 3 se ve que el coeficiente acotado tiene una oscilación menor y simétrica en relación con la media; mientras el coeficiente de Pearson presenta un rango mayor y decreciente a medida que aumenta la media.

Para una mejor comprensión interpretativa de estos estadísticos, se presentan 8 hipotéticas distribuciones de la ideología con sus correspondientes medias, desviaciones típicas y porcentajes de desviación empírica (tabla 4). En los dos primeros lugares, tendríamos la situación en la que todos los sujetos de nuestro estudio tuviesen la misma ideología (4 en la distribución A y 3 en la distribución B); en ambos casos la varianza sería 0 y, por tanto, aunque la desviación máxima —es decir, el denominador de la última fórmula— fuese de 9 u 8, el PV_a seguirá siendo de 0. Ahora bien, si la media de la variable es 4 y los sujetos están ubicados —sólo y necesariamente— a partes iguales en la extrema izquierda y en la extrema derecha, entonces la varianza es máxima y, por tanto, nuestro coeficiente es del cien por cien. Una segunda situación que debiera arrojar la misma cantidad de porcentaje de desviación empírica sería aquella en la que no siendo la media igual a 4, sólo existen valores extremos (distribución D) En tal caso, también el porcentaje de variación es del cien por cien: todos lo sujetos están en los extremos de la distribución.

Situaciones menos extremas se presentan en las cuatro siguientes distribuciones: en la E y en la F, hemos generado sendas tablas en las que se han situado el 50 por 100 de los sujetos en la media y otros tantos —repartidos en partes iguales— en los valores extremos. En esos casos, el porcentaje de variación es del 50 por 100. En cambio, si además de los casos extremos, existen valores intermedios, este nuevo coeficiente disminuirá en proporción siempre inferior al número de sujetos que no tengan valores extremos. Comparándola con la distribución E, en la distribución G se contempla cómo habiendo salido de los extremos un 38 por 100 de casos, el coeficiente sólo disminuye 21 puntos porcentuales y, del mismo modo, la distribución H tiene un 25 por 100 de variación y, sin embargo, sólo un 50 por 100 de los casos están ubicados en la media y un 18 por 100 tienen valores extremos.

Valores de los coeficientes en los modelos de probabilidad

Una importante pregunta para el uso e interpretación de estos coeficientes sería la de ver cómo se comportan cuando se aplican a modelos de distribuciones de probabilidad. Véanse tres de las distribuciones de probabilidad más importantes en la estadística: la constante, la binomial y la normal. Empecemos por las dos primeras, pues ambas son distribuciones acotadas.

Por definición, una distribución uniforme es aquella en la que todos los valores poseen la misma probabilidad de salir. Está definida por dos parámetros que son *a* y *b*, que son, respectivamente, los valores mínimo y máximo de la distribución. La función de densidad de esta variable es:

$$f(x) = \frac{1}{b-a} \qquad [8]$$

TABLA 4

Distribuciones simuladas de ideología

Distribución A	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$	*Distribución B*	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$
Centro	4	1,0	4,0	0,0	Cent.-izquierda	3	1,0	3,0	0,0
			$\bar{x}$ = 4,0	s^2 = 0,0				$\bar{x}$ = 3,0	s^2 = 0,0
PV_a = 0%					PV_a = 0%				
CV_a = 0%					CV_a = 0%				

Distribución C	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$	*Distribución D*	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$
Ext. izquierda	1	0,5	0,5	4,5	Ext. izquierda	1	0,7	0,7	2,7
Ext. derecha	7	0,5	3,5	4,5	Ext. derecha	7	0,3	2,3	5,3
			$\bar{x}$ = 4,0	s^2 = 9,0				$\bar{x}$ = 3,0	s^2 = 8,0
PV_a = 100%					PV_a = 100%				
CV_a = 100%					CV_a = 100%				

Distribución E	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$	*Distribución F*	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$
Ext. izquierda	1	0,3	0,3	2,3	Ext izquierda	1	0,3	0,3	1,3
Centro	4	0,5	2,0	0,0	Cent.-izquierda	3	0,5	1,5	0,0
Ext. derecha	7	0,3	1,8	2,3	Ext. derecha	7	0,2	1,2	2,7
			$\bar{x}$ = 4,0	s^2 = 4,5				$\bar{x}$ = 3,0	s^2 = 4,0
PV_a = 50%					PV_a = 50%				
CV_a = 71%					CV_a = 71%				

Distribución G	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$	*Distribución H*	x_i	p_i	x_ip_i	$(x_i-\bar{x})^2p_i$
Ext. izquierda	1	0,0	0,0	0,2	Ext. izquierda	1	0,1	0,1	0,2
Cent.-izquierda	2	0,2	0,5	0,9	Izquierda	2	0,3	0,6	0,3
Centro	4	0,5	2,0	0,0	Cent.-izquierda	3	0,5	1,5	0,0
Cent.-derecha	6	0,2	1,4	0,9	Centro	6	0,1	0,5	0,8
Ext. derecha	7	0,0	0,2	0,2	Ext. derecha	7	0,0	0,3	0,6
			$\bar{x}$ = 4,0	s^2 = 4,5				$\bar{x}$ = 3,0	s^2 = 4,0
PV_a = 25%					PV_a = 25%				
CV_a = 50%					CV_a = 50%				

Además, se sabe que la media y la varianza de esta distribución presentan necesariamente los siguientes valores:

$$E(X) = \frac{a+b}{2} \qquad [9]$$

$$Var(X) = \frac{(b-a)^2}{12}$$

Conociendo que la media viene dada por $(a + b) / 2$, la varianza máxima será

$$Var_{max} = \frac{(b-a)^2}{4} \qquad [10]$$

Por consiguiente, la proporción y el coeficiente de variación acotada tendrán los siguientes valores:

$$PV_a = \frac{1}{3} = 33{,}3\% \qquad [11]$$

$$CV_a = \sqrt{\frac{1}{3}} = 57{,}7\%$$

En el caso de la distribución binomial, que posee como función de distribución la fórmula:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad [12]$$

Los valores máximo y mínimo de esta distribución son, respectivamente, 0 y n, y la esperanza y varianza se ajustan a las siguientes expresiones:

$$E(X) = np \qquad [13]$$

$$Var(X) = np(1-p)$$

Como quiera que la varianza máxima adquiere el siguiente valor:

$$Var_{max} = (np - 0)\ (n - np) - n^2 p(1-p) \qquad [14]$$

Los valores correspondientes a la proporción y coeficientes acotados estarán en función del número de veces que se repita el experimento binomial:

$$PV_a = \frac{npq}{n^2pq} = \frac{1}{n} \qquad [15]$$

$$CV_a = \sqrt{\frac{1}{n}}$$

Por último, véase la distribución normal, cuya función de distribución para variables tipificadas ($\mu = 0$ y $\sigma = 1$) es la siguiente:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad [16]$$

El cálculo del coeficiente de variación acotada para la distribución normal plantea un no leve problema, porque esta distribución no tiene límites inferior ni superior. Sin embargo, con el propósito de tener una referencia de este coeficiente con la distribución normal, se puede aceptar el supuesto de que los extremos de la distribución están situados a más/menos tres desviaciones típicas. Este supuesto sólo deja de ser cierto para menos del 0,3 por 100 de los casos, por lo que su peso ha de tener escasa influencia sobre el estadístico hallado. De este modo, la proporción de varianza acotada en la distribución normal sería del 11 por 100 y el coeficiente de variación del 33 por 100.

$$CV_a = \sqrt{\frac{1}{(0-(-3))(3-0)}} = \sqrt{\frac{1}{9}} = 33\% \qquad [17]$$

Otros coeficientes de diversidad

Evidentemente, el aquí propuesto no es el único coeficiente de diversidad que varía entre 0 y 1. Existen otros en la literatura estadística que es conveniente presentar con el fin de compararlos y juzgar las bondades de unos y otros.

En primer lugar, es preciso sacar a colación el *índice de Gini*. Éste es un índice de concentración de distribuciones, muy utilizado para medir la desigualdad económica. Para su cálculo existen dos fórmulas: una que consiste en el sumatorio de las diferencias entre porcentajes de cantidades ($x_i \times p_i$) y porcentajes de casos (p_i); la otra consiste en sumar las diferencias entre cada valor con el resto. Entre ellas, además, se da una curiosa relación tautológico-matemática.

$$G = \frac{\sum_{i=l}^{k-l} P_i - Q_i}{\sum_{i=l}^{k-l} P_i} \qquad [18]$$

$$G = \frac{\sum_{i=l}^{n} \sum_{j=l}^{n} |X_i - X_j|}{2\overline{X}(n-1)n}, \; i \neq j$$

Siempre y cuando el valor mínimo de la variable sea nulo, el rango de este coeficiente también varía entre 0 y 1, si bien sus resultados dan cuenta de otro tipo de desviación en las distribuciones. El valor 0 se obtiene en el caso de que $P_i = Q_i; \forall_i$, es decir, siempre que haya distribución igualitaria: todos los sujetos tienen el mismo valor. De esta forma coincide con los coeficientes de variación puesto que en tales condiciones siempre arrojan el valor 0. Ahora bien, sólo dan el máximo valor (1) si sólo existen dos valores distintos y el mínimo es 0. Esto le diferencia de la proporción de varianza, puesto que para alcanzar su valor máximo, ésta no requiere que el valor mínimo sea 0. Por otro lado, el coeficiente de Gini tiene mal ajuste cuando los valores de las variables son negativos. Además, tiene la misma propiedad que el coeficiente de variación de Pearson: al multiplicar por una constante los valores de la variable, se mantiene inalterado el coeficiente; pero no sucede lo mismo si se le añade una determinada cantidad.

Del mismo modo que se ha procedido con la varianza y la desviación típica, podríamos operar con el índice de Gini. Se acaba de mencionar que para obtener su valor máximo (1) en una distribución es condición necesaria que el valor mínimo de la variable sea el 0, porque si la distribución empezara con un valor distinto, el coeficiente máximo posible tendría la siguiente expresión[3]:

$$G_{max} = 1 - \frac{X_{min}}{\overline{X}} \qquad [19]$$

Conociendo de este modo el valor máximo que puede adoptar este coeficiente en el supuesto de que el valor mínimo sea distinto de 0, se puede obtener una medida ajustada que responde al siguiente cociente:

$$G_a = \frac{G}{G_{max}} \qquad [20]$$

que posee las dos siguientes propiedades adicionales:

a) Varía entre 0 y 1.
b) La adición de una constante a la variable no modifica el valor del coeficiente.

Otro índice de dispersión que también varía entre 0 y 1 es el *índice de variación cualitativa.* Su fórmula viene dada por la siguiente expresión:

[3] Suponemos que los valores de las variables son positivos.

$$IVC = \frac{(1 - \sum_{i}^{k} p_i^2)}{(k-1)/k} \qquad [21]$$

Este índice, aplicable a variables no métricas, muestra el grado de concentración de los sujetos en determinadas categorías de la variable, por lo que no da cuenta de la distancia entre los valores. El coeficiente es cero cuando el cien por cien de los casos se concentra en un solo valor de la variable. En cambio, es 1 cuando las frecuencias de todos los valores de la variable son iguales. El problema con este índice es que no tiene una correcta aplicación en el caso de variables en las que importe la distancia entre valores. Quiero decir con ello que en una variable con valores muy de acuerdo, de acuerdo, en desacuerdo y muy en desacuerdo, el resultado sería el mismo si sólo hay casos en el muy de acuerdo y en el muy en desacuerdo, que si sólo hay respuesta con muy de acuerdo y de acuerdo. Y es evidente que en el primer caso la dispersión debería ser mayor.

Otro coeficiente aplicable a variables nominales es el de *entropía,* estadístico que procede de la teoría de la información (Kripendorff, 1986). Es una medida que da cuenta de la cantidad de incertidumbre que proporciona una variable. Nula, igual a 0, cuando todos los casos caen en una sola categoría; total incertidumbre, cuando los casos se reparten a partes iguales entre las categorías de las variables. Su fórmula viene dada por la siguiente expresión:

$$H = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \qquad [22]$$

Este coeficiente tiene un rango que va desde 0 a $\log_2(n)$, por lo que si lo dividimos por esta cantidad se puede obtener un coeficiente normalizado con límites situados entre 0 y 1; pero, a semejanza del índice de variación cualitativa, este cociente es insensible a los valores que pueden adoptar las variables y, por tanto, es aplicable a variables nominales y no a variables cuantitativas acotadas, que es el caso que aquí nos ocupa.

Sin embargo, existe en la literatura estadística otro coeficiente, muy similar al de la entropía y muy apropiado para el estudio de la desigualdad de las distribuciones. Se trata del *índice de Theil.* La conexión entre uno y otro es la siguiente: en primer lugar, en vez de construir el índice de entropía con las proporciones de la variable, se calcula a partir de las proporciones de las cantidades ($x_i \times p_i$ o q_i):

$$H = \sum_{i=1}^{n} q_i \log_2 \frac{1}{q_i} \qquad [23]$$

El resultado de esta fórmula, que se encuentra en un rango entre 0 y el logaritmo de *n*, debe ser invertido puesto que es tanto mayor cuanto menos equitativamente se distribuye la variable. Por tanto, el índice de Theil (H') será:

$$H' = \log_2 n - H \qquad [24]$$

$$= \sum_{i=1}^{n} q_i \log_2 \frac{q_i}{p_i}$$

Lógicamente, en consecuencia, el nuevo índice tendrá un rango entre 0 y $\log_2 n$. Por ello, para su normalización los expertos proponen que se divida por tal logaritmo.

$$H_{\bar{N}} = \frac{H'}{\log_2 n} \qquad [25]$$

Sin embargo, en paralelismo con el ya expuesto coeficiente de variación acotada, en las variables con valores limitados, el valor máximo que puede adquirir H', dada una media $\bar{X}$, es

$$H'_{max} = \frac{X_{min}(X_{max} - \bar{X}) \log_2 \frac{X_{min}}{\bar{X}} + X_{max}(\bar{X} - X_{min}) \log_2 \frac{X_{max}}{\bar{X}}}{\bar{X}(X_{max} - X_{min})} \qquad [26]$$

A partir de ésta, puede hallarse un coeficiente que dé un resultado entre 0 y 1, conocidos el valor medio, el mínimo y el máximo: basta con obtener el cociente entre el índice de Theil y el valor máximo representado por H'_{max}. A este nuevo coeficiente le llamaremos *índice de Theil acotado.*

$$H'_a = \frac{H'}{H'_{max}} \qquad [27]$$

Veamos un ejemplo real, pero no complejo, donde se muestre la utilidad de esta proporción utilizada en lugar del coeficiente de entropía típico (normalizado). Con los datos del Censo de 1991, tomando como base las mujeres censadas en la ciudad de Salamanca con más de 15 años, se ha obtenido la distribución en el número de hijos presentada en la tabla 5.

TABLA 5

Distribución del número de hijos de mujeres de más de 15 años (Salamanca)

	xi	fi			Índices Acotados
Salamanca	0	29497	Media:	1,7	
Mujeres >15	1	7861	Varianza:	3,6	34,3%
	2	14379	DT:	1,9	58,6%
	3	8119	R_m:	2,3	34,3%
	4	4871	Theil:	0,69	44,2%
	5	2943	Gini:	0,39	38,9%
	6	1624	DM Gini:	1,3	38,9%
	8	1624	Entropía:	1,7	14,9%
		70917	Asimetría:	1,2	
			Curtosis:	1,3	
			DM:	0,8	30,6%

El número medio de hijos por mujer salmantina es de 1,7. Con esta media, un valor mínimo de 0 y un valor máximo de 8[4], el H'_a adopta el valor 0,44. Quiere ello decir que el índice de Theil adopta un valor que es el 44 por 100 del máximo posible dadas las características de la distribución. Si se hubiera calculado el índice de Theil normalizado, el resultado hubiese sido de 0,02. Un examen atento de la distribución nos inclina a decantarnos por la primera medida en lugar de la segunda.

Para terminar con este repaso de estadísticos de dispersión, vamos a hacer referencia al *promedio de las razones de ventaja*. Este último es el cociente entre la cantidad repartida (q_i) y la población afectada por la repartición (p_i) para cada uno de los escalones ($_i$) de una distribución. Dicho con otras palabras, es el índice de Theil, pero calculado sin logaritmos:

$$R_m = \sum_{i=1}^{n} q_i \frac{q_i}{p_i} \qquad [28]$$

Este coeficiente varía entre 1 (valor mínimo, cuando la variable se reparte por igual entre todos los casos) y *n*, en el caso de que una unidad se adueñe del sumatorio total de la variable. Por ello, se ha propuesto como coeficiente normalizado del promedio de razones la siguiente expresión:

[4] La última fila de la tabla de distribución de frecuencias representa el valor tener 7 hijos o más; el valor 8 sería la marca de clase de este intervalo. El cambio de este valor dentro de límites razonables afectaría mínimamente a la proporción de Theil acotada. Así, con valor 7 H'max = 2,09; con 8, 2,27; con 9, 2,42; con 10, 2,55; con 11, 2,67. Además, suponiendo marcas de clase entre 7 y 11, el PH varía sólo entre 0,14 y 0,15.

$$R_m^N = \frac{R_m - 1}{n - 1} \qquad [29]$$

Ahora bien, este «adueñamiento» es imposible en variables acotadas por la misma razón que vimos con el coeficiente de variación (acuérdense de Matusalén). Por ello, aquí se propone también una medida que utilice en el cociente el valor máximo de este coeficiente dados una media, un valor mínimo y un valor máximo de la distribución. No es difícil obtener con un camino similar hasta el ahora realizado el coeficiente máximo del promedio de razones de ventaja. Éste es:

$$R_{m_{\max}} = \frac{(\bar{X}X_{\min}) + (\bar{X}X_{\max}) - (X_{\max}X_{\min})}{\bar{X}^2} \qquad [30]$$

Consecuentemente, podría calcularse la proporción del promedio de las razones de ventaja con el cociente de las dos últimas fórmulas:

$$\frac{R_m}{R_{m_{\max}}} \qquad [31]$$

Sin embargo, debido a que el límite inferior de ambos cocientes es 1, mejoraría la proporción si con la conveniente resta bajamos el punto mínimo de la escala de ambas medidas a 0:

$$RM_a = \frac{R_m - 1}{R_{m_{\max}} - 1} \qquad [32]$$

Y, sorprendentemente, se da en todos los casos la relación algebraica siguiente:

$$RM_a = PV_a \qquad [33]$$

Aplicaciones

Para terminar este artículo, se comentan a continuación dos ejemplos que ponen a prueba la utilidad y la conveniencia de los coeficientes propuestos.

El primer ejemplo tiene su campo de aplicación en la desigualdad educativa, donde las medidas de diversidad arrojan en muchas ocasiones resultados contradictorios que hay que saber explicar y, sobre todo, nos advierten de que si no empleamos la correcta medida podemos llegar a conclusiones erróneas.

Obtenida a partir de la encuesta sociodemográfica, que por su tamaño muestral nos permite estudiar la desigualdad educativa por cohortes sin problemas de errores muestrales, la tabla 6, con la debida precaución por los pro-

TABLA 6

Número de años escolarizados según cohorte y género

	Todos				*Varones*				*Mujeres*			
Cohorte	*Media*	*t.DT*	*t.CV (%)*	*t.CVa (%)*	*Media*	*v.DT*	*v.CV (%)*	*v.CVa (%)*	*Media*	*m.DT*	*m.CV (%)*	*m.CVa (%)*
1962-1966 (25 a 29 años)	9,2	3,4	37,0	34,1	8,9	3,3	37,1	33,2	9,5	3,6	37,9	36,0
1957-1961 (30 a 34 años)	8,0	4,2	52,2	42,8	8,0	4,1	51,3	41,8	8,1	4,3	53,1	43,8
1952-1956 (35 a 39 años)	7,1	4,2	59,2	43,9	7,3	4,2	57,5	43,6	6,9	4,2	60,9	44,2
1947-1951 (40 a 44 años)	6,4	4,0	62,5	42,9	6,5	4,1	63,1	43,8	6,2	3,8	61,3	41,1
1942-1946 (45 a 49 años)	5,8	3,9	67,2	43,0	6,1	4,1	67,2	44,5	5,5	3,7	67,3	41,4
1937-1941 (50 a 54 años)	4,9	3,6	73,5	41,9	5,2	3,7	71,2	42,2	4,6	3,3	71,7	39,2
1932-1936 (55 a 59 años)	4,5	3,5	77,8	41,9	4,8	3,6	75,0	42,1	4,2	3,3	78,6	40,5
1927-1931 (60 a 64 años)	4,2	3,2	76,2	39,3	4,4	3,4	77,3	41,0	4,0	3,0	75,0	37,5
1922-1926 (65 a 69 años)	4,1	3,2	78,0	39,6	4,3	3,3	76,7	40,2	3,9	3,0	76,9	37,9
1917-1921 (70 a 74 años)	3,9	3,1	79,5	39,1	4,3	3,3	76,7	40,2	3,5	2,9	82,9	38,2
1916 y antes (75 y más)	3,5	3,1	88,6	40,8	3,9	3,2	82,1	40,4	3,1	2,8	90,3	38,7

FUENTE: INE, *Encuesta Sociodemográfica* (1991). Elaboracion propia.

blemas que pudiera ocasionar la mortalidad diferencial, muestra con claridad que entre la generación anterior al 17 y la generación de inicios del 60 ha habido un aumento considerable de los años de escolarización entre los españoles: de tres años y medio a algo más de nueve años. Este aumento ha sido más espectacular entre las mujeres, que han pasado de 3 años a 9,5, que entre los varones, que de 3,9 saltan a menos de 9. Dicho con otras palabras, antes de la generación del 56 los hombres tenían una escolarización mayor que las mujeres; a partir de entonces son las mujeres quienes más tiempo permanecen en el sistema educativo.

La desviación típica aumenta sistemáticamente excepto en la última cohorte, cuyos datos hay que tomar con precaución porque en ella aún están personas de 25 a 29 años que todavía no han podido terminar su licenciatura o doctorado, por lo que de ella se espera un aumento de la media y de la desviación típica. Pues bien, este constante aumento de la desviación típica proviene principalmente del aumento del porcentaje de personas que han venido cursando los estudios superiores y doctorado: en la medida que hay un aumento de la media por debajo del punto medio de una escala (aquí estamos tratando con una con unos límites entre 0 y 20; por tanto, el punto medio es 10), la desviación típica va aumentando.

Pero resalta en el coeficiente de variación un drástico descenso de la desviación educativa: desde un 88,6 por 100 para la cohorte con más de 75 años hasta el —seguramente infravalorado— 37 por 100 de la cohorte de los nacidos entre el 62 y el 66. Como la media aumenta bastante más que lo hace la desviación típica, es obvio que el coeficiente de variación descienda. Sin embargo, ¿puede afirmarse que la desigualdad educativa haya descendido con tanta intensidad y tan regularmente? ¿En qué puede confiarse más, en la desviación típica o en el coeficiente de variación? La respuesta a esta última pregunta es obvia: es preferible utilizar un coeficiente relativo a otro absoluto para realizar comparaciones; pero también se puede hablar con propiedad diciendo que un coeficiente mide desviación absoluta y el otro relativa y según lo que queramos comparar se debería utilizar una medida u otra.

Sin embargo, si se tiene presente el coeficiente de variación acotada que se propone en este artículo, los resultados son diferentes a los dos anteriores: muestran una determinada estabilidad en la desigualdad educativa a través de las cohortes. Y vienen a mostrar que, haciendo abstracción de la media de escolarización, la desigualdad educativa no disminuye durante los primeros años del franquismo, sino que más bien aumenta en relación con la de generaciones anteriores.

Otro síntoma que avala el coeficiente de variación acotada sería la comparación entre hombres y mujeres: hasta la generación del 52 la variación educativa era mayor entre los hombres que entre las mujeres. Como la mayor parte de ellas se quedaban en los niveles bajos del sistema educativo, la desigualdad era baja. Es a partir de los años sesenta, con la incipiente incorporación a los niveles medios y superiores de las jóvenes de clase media-alta, cuando empiezan a aumentar las desigualdades entre ellas por encima de las de los hombres.

TABLA 7

Número de años de escolarización. España, 1964-1992

Año	*Media*	*DT*	*CVa (%)*	*CV (%)*
1964	4,874	2,698	31	55
1965	4,906	2,712	32	55
1966	4,941	2,729	32	55
1967	4,979	2,747	32	55
1968	5,041	2,780	32	55
1969	5,108	2,828	32	55
1970	5,189	2,898	33	56
1971	5,283	2,983	34	56
1972	5,373	3,042	34	57
1973	5,467	3,099	35	57
1974	5,564	3,160	35	57
1975	5,645	3,206	36	57
1976	5,752	3,262	36	57
1977	5,807	3,292	36	57
1978	5,898	3,348	37	57
1979	5,933	3,349	37	56
1980	6,005	3,432	37	57
1981	6,099	3,465	38	57
1982	6,230	3,509	38	56
1983	6,378	3,600	39	56
1984	6,469	3,624	39	56
1985	6,618	3,674	39	56
1986	6,718	3,708	39	55
1987	6,809	3,749	40	55
1988	6,906	3,783	40	55
1989	7,054	3,851	40	55
1990	7,155	3,874	40	54
1991	7,250	3,893	40	54
1992	7,346	3,917	41	53

FUENTE: Mas, Pérez, Uriel y Serrano (1995).

Se podría objetar que el hecho de que el coeficiente de variación acotada no varíe entre las cohortes es un síntoma de su poca sensibilidad a las desviaciones y que por ello es una medida poco útil. Para demostrar que ello no es así, voy a enseñar unos datos de la misma variable; pero en lugar de ser presentados por cohortes, están presentados de modo transversal aunque en forma de serie temporal (tabla 7). Son unos datos procedentes de la recopilación efectuada por Mas, Uriel y Serrano. Para cada año se proporciona media, desviación típica, coeficiente de variación acotada y coeficiente de variación de Pearson del conjunto de la población. Aquí es obvio esperar que la desviación educativa vaya aumentando porque las nuevas generaciones que se incorporan lo hacen con un número medio de años escolarizados muy superior a la media y se aprecia fácilmente cómo el coeficiente de variación acotada pasa del 31 al 41 por 100. Además, la transición es creciente y monótona entre los 29 años relatados. En ningún momento el coeficiente aumenta y nunca disminuye más de dos puntos porcentuales.

Siguiendo con la aplicación educativa, resulta en este contexto interesante traer a colación los datos que Ram publicó en 1990 en la *Review of Economics and Statistics,* en la que relacionaba las medias y las desviaciones típicas de la escolarización en 94 países distintos. En el gráfico 4 se observa claramente la relación curvilínea existente entre las medias —en el eje de abcisas— y las desviaciones —en el eje de ordenadas representadas por rectángulos—. Esta relación empírica, que Ram expresa en una ecuación de regresión cuadrática, no debe sorprender después de haber visto el comportamiento que tiene la desviación típica en variables acotadas. Su valor disminuye tautológicamente cuando la media se encuentra en los valores extremos de la distribución. La conclusión que se puede extraer de estos datos puede ser ficticia: decir que al aumentar el nivel educativo medio en países en desarrollo aumenta las desigualdades educativas es una verdad a medias. Aunque inicialmente las aumente, al final se verán disminuidas cuando la media de la escolarización supere el punto medio de la escala.

Pero aún hay más: en términos de desviación relativa, los resultados serían distintos pues, como es apreciable en el gráfico citado, donde se representan medias y coeficiente de variación de los mismos 94 países, la desigualdad (relativa) va disminuyendo drásticamente en relación inversa a medida que aumenta la media. De la misma forma, si se observa en el mismo gráfico con el coeficiente de variación acotada se aprecia que la desigualdad educativa no sólo no aumenta a medida que se acrecienta la media, sino más bien al contrario, encontrando una tenue relación lineal inversa, no tan pronunciada en sus inicios por el coeficiente de variación, por el hecho de estar éste sesgado hacia lo alto cuando la media es próxima a 0.

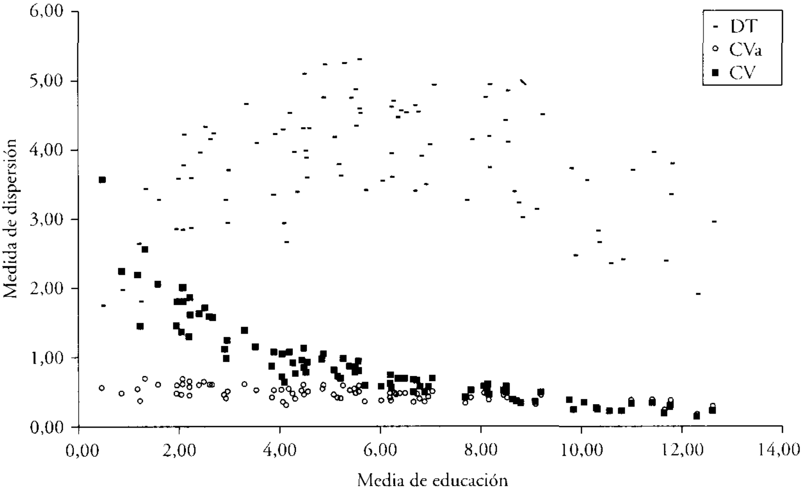
El segundo ejemplo también es longitudinal, abarca desde junio del 79 hasta junio del 85 y está sacado de los distintos estudios del CIS. En él se emplean 29 estudios en los que se preguntó por la ideología de los sujetos. En los quince primeros (hasta junio del 82) se utiliza una escala del 1 al 7; en los catorce últimos se emplea una escala del 1 al 10. Veamos los efectos que ello tiene tanto sobre la media como sobre los coeficientes de variación.

Para el estudio de la evolución de la media se presenta el gráfico 5. En él la línea gruesa representa la ideología medida antes de julio del 83 con la escala de 7 puntos. Con la línea delgada, la ideología medida a partir de la mencionada fecha, pero con la escala de 10 puntos. También se ha representado una línea punteada que representa la media obtenida antes del 83 en el caso de que se hubiese medido con la escala de 10 puntos[5].

[5] Para ello, se le ha aplicado una transformación lineal de modo que los extremos de la una coincidiesen con los de la otra. Los parámetros *a* y *b* de esa transformación son –0,5 y 1,5, respectivamente.

GRÁFICO 4

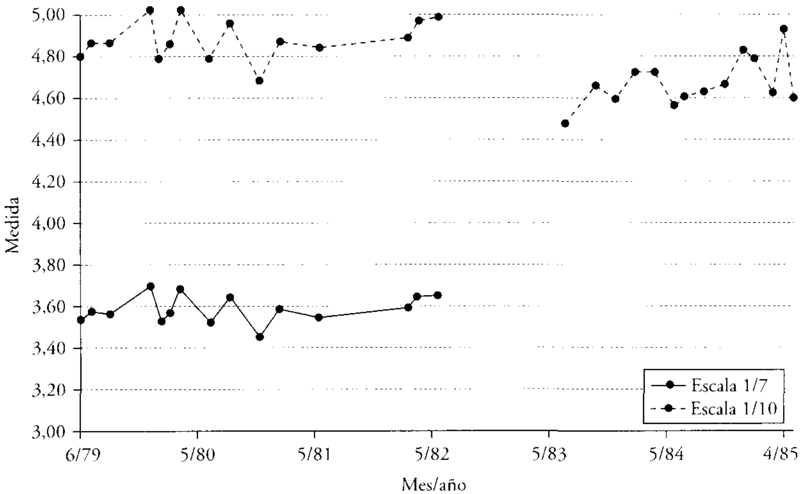
Medias de dispersión según media de años de educación en 94 países



FUENTE: Ram (1990).

GRÁFICO 5

Evolución de la media de la escala de ideología

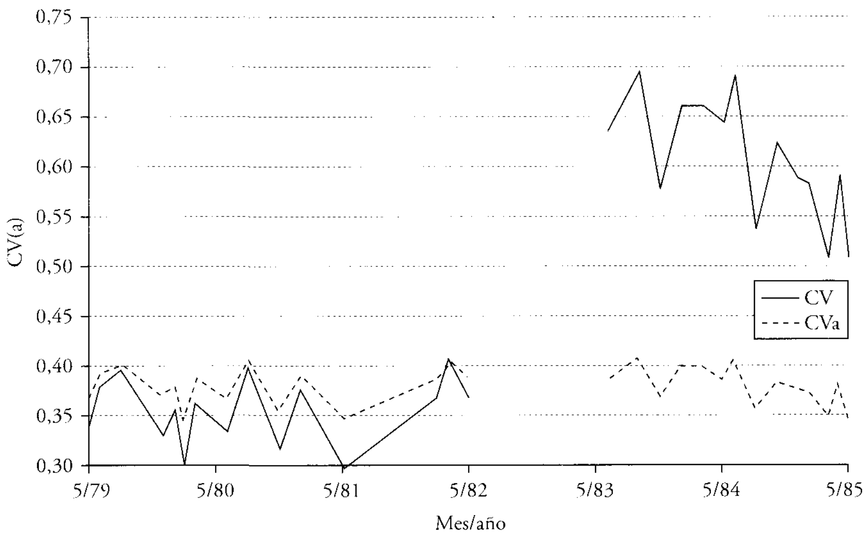


FUENTE: CIS (1979-1985).

Como era de esperar, entre la línea gruesa y la delgada hay un salto ficticio producido por el alargamiento artificial de la escala. Sin embargo, al comparar la línea punteada con la fina, se observa un ligero decaimiento hacia abajo de la media. Esto, aunque se sale de la línea argumental de este artículo, se puede explicar de dos maneras: una, porque el gobierno del partido socialista hizo que la población se identificase más a la izquierda, y otra, porque el haber puesto una escala del 1 al 10 provoca el efecto de confundir el centro (5) con el centro-izquierda, pues el verdadero centro de la escala (5,5) es imposible de declarar. Lo que más importa aquí es ver qué es lo que pasa con la dispersión de la distribución. La hipótesis de partida es que ni el cambio de situación política ni el cambio de la escala tendría que repercutir en la polarización ideológica de la población. El gráfico 6, en el que se comparan los coeficientes de variación objeto de esta presentación, muestra un comportamiento muy distinto en uno y en otro. El coeficiente de variación de Pearson es sensible al cambio de escala: de estar en el intervalo 0,30-0,40 hasta el 82, pasa al intervalo entre 0,50-0,70 a partir del 83. En cambio, el coeficiente de variación acotada se mantiene en el intervalo 0,35-0,40 durante los dos períodos. Además, en el segundo período los dos coeficientes muestran unos movimientos similares, aunque de mayor amplitud en el caso del de Pearson, que por no aparecer en el período anterior son sujetos de sospecha.

GRÁFICO 6

Evolución de los coeficientes de variación de la escala de ideología



FUENTE: CIS (1979-1985).

En definitiva, todos los indicios muestran que con el objeto de comparar distribuciones distintas con escalas diferentes parecen más adecuados los coeficientes de variación acotada aquí propuestos. Partiendo de la varianza, se ha pasado revista a una serie de estadísticos conocidos como el coeficiente de Gini, los índices de Theil y la razón media, se ha estudiado cuáles son sus límites para utilizarlos como denominador y de este modo se ha obtenido una serie de medidas ajustadas con límites verdaderos entre 0 y 1, propiedad de la que carecen los coeficientes normalizados conocidos en la literatura estadística.

La idea, generalizable a toda una serie de estadísticos importantes para el estudio de la desviación, la desigualdad y polarización, se ha mostrado congruente, hasta el punto de que al aplicarla a dos estadísticos distintos proporciona el mismo resultado y, a través de dos aplicaciones distintas —desigualdad educativa y polarización ideológica—, se ha visto las ventajas prácticas de su utilización. Aún queda investigación pendiente. Por ejemplo, el cálculo del error típico de estos estadísticos y un estudio más sosegado sobre la función de bienestar que incorporan, porque, como muy bien dijo Atkinson, en toda medida de desigualdad hay implícita una concepción de la justicia.

BIBLIOGRAFÍA

ATKINSON, A. B. (1970): On the Measurement of Inequality, *Journal of Economic Theory*, 2: 244-263.

BLALOCK, H. M. (1991): *Understanding Social Inequality*, Newbury Park: Sage.

BOSCH, A.; ESCRIBANO, C., y SÁNCHEZ, I. (1989): *Evolución de la desigualdad y la pobreza en España. Estudio basado en las Encuestas de Presupuestos Familiares 1973-1974 y 1980-1981*, Madrid: INE.

BOSSERT, W., y PFINGSTEN, A. (1990): «Intermediate Inequality: Concepts, Indices, and Welfare Implications», *Mathematical Social Sciences*, 19: 117-134.

CORTÉS, F., y RUBALCABA, R. M. (1984): *Técnicas estadísticas para el estudio de la desigualdad social*, México: El Colegio de México.

COULTER, P. B. (1984): «Distinguishing Inequality and Concentration», *Political Methodology*, 10: 323-335.

ESTEBAN, J. M., y RAY, D. (1993): «El concepto de polarización y su medición», *I Simposio sobre Igualdad y Distribución de la Renta y Riqueza*, Madrid: Fundación Argentaria.

LEABO, D. A. (1976): *Basic Statistics*, Homewood: Irwin.

KENDALL, M. G., y STUART, A. (1973): *The Advanced Theory of Statistics*, London: Griffin.

KOLM, S. C. (1976*a*): «Unequal Inequalities I», *Journal of Economic Theory*, 12: 416-442.

— (1976*b*): «Unequal Inequalities II», *Journal of Economic Theory*, 13: 82-111.

KOTZ, S.; JOHNSON, N. L., y READ, C. B. (1983): *Encyclopedia of Statistical Science*, 8 vols., New York: John Wiley.

KRIPPENDORF, K. (1986): *Information Theory*, Newbury Park: Sage.

MAS, M.; PÉREZ, F.; URIEL, E., y SERRANO, L. (1995): *Capital Humano. Series Históricas, 1964-1992*, Fundación Bancaja.

RAM, R. (1990): «Educational Expansion and Scholing Inequality: International Evidence and Some Implications», *The Review of Economics and Statistics*, 266-273.

RAYNER, S. C. W. (1975): «Variance Bounds», *Indian Journal of Statistics*, 37: 135-138.

RUIZ CASTILLO, J. (1993): «Distribución personal de la renta: medición empírica y juicios de valor», *I Simposio sobre Igualdad y Distribución de la Renta y Riqueza,* Madrid: Fundación Argentaria.

THEIL, H. (1967): *Economics and Information Theory,* Amsterdam: North Holland Publishing Co.

WALDMAN, L. K. (1976): «Measures of party systems' properties: The number and sizes of parties», *Political Methodology,* 3: 199-214.

WEISBERG, H. F. (1986): *Central Tendency and Variability,* Newbury Park: Sage.

ABSTRACT

This paper distinguishes between the concepts of deviation, inequality and polarisation and proposes a set of coefficients which can be applied to variables with limited —ie, minimum and maximum— values, for the purpose of measuring the deviation of distributions in percentile terms. Of all the coefficients proposed by the author, the simplest and most useful is the adjusted variation coefficient which implies adding a series of desired properties to the well-known Pearson correlation coefficient for measuring limited variables. The author also presents a pair of applications which highlight the advantages of this coefficient.