

Dirección General de Estadísticas Sociodemográficas

Nota Técnica

Interpretación del coeficiente de variación, y su relación
con el error relativo de las estimaciones

INSTITUTO NACIONAL DE GEOGRAFIA Y ESTADISTICA

5 de diciembre de 2017

Introducción

El coeficiente de variación (CV) de una estimación y su error relativo, son dos índices de precisión útiles para el análisis de las encuestas y son, junto con el nivel de confiabilidad deseado y el error estándar de la estimación, muy útiles para evaluar la calidad de las estimaciones obtenidas. Como veremos en el documento, existe una relación directa entre el CV y el error relativo, lo cual nos podría ayudar enormemente a clarificar las cotas en los valores del CV que aún están en discusión.

Como parte del compromiso y profesionalismo que tenemos ante la sociedad, la DGES realizó este diagnóstico sobre el CV con el fin de fijar una postura acerca de la preocupación que se tiene sobre el establecimiento institucional de grandes cotas a los coeficientes de variación.

Es necesario considerar que el coeficiente de variación no es el único indicador de la precisión de las estimaciones, sin embargo, sí es el único indicador para el cual se proporciona una interpretación institucional. De hecho, siempre será recomendable utilizar el resto de indicadores de precisión que permitan evaluar la calidad de las estimaciones.

Decir al usuario que una estimación es confiable cuando en realidad tiene un rango de posibles valores bastante amplio, es bastante peligroso. Recordemos que la Ley del SNIEG obliga a las unidades del estado a usar la información que producimos; es por eso que debemos garantizar que la información que les proporcionamos, les servirá para tomar buenas decisiones.

La postura de las encuestas en hogares, ha sido mantenerse en los cortes de 15% y 25% en el valor del coeficiente de variación dado que, como se verá en el presente documento, hay una relación directa entre el error relativo y el coeficiente de variación.

Objetivo

El objetivo del presente documento es mostrar la importancia que tiene la relación entre el coeficiente de variación de una estimación y su error relativo. Como se mencionó anteriormente, esto nos permite establecer el rango esperado del intervalo, lo que a su vez permite a las personas que toman decisiones, establecer el nivel de confianza en las estimaciones que proporcionan las muestras. Finalmente, lograr establecer una postura institucional sobre el coeficiente de variación, dado que actualmente está diferenciado en las distintas áreas que integran el instituto.

Definición de coeficiente de variación

El coeficiente de variación es una medida de dispersión que se usa para evaluar la calidad de las estimaciones. Se define como:

$$cv_{\mu} = \frac{\sigma_{\mu}}{|\mu|}$$

donde:

cv_{μ} es el coeficiente de variación del estimador μ

σ_{μ} es el error estándar del estimador μ

μ estimador de interés.

Cuando se analizan los resultados de una muestra, se utilizan los valores estimados de la media y el error estándar anteriormente descritos, y con ellos generamos una estimación del coeficiente de variación.

Es importante observar que el coeficiente de variación es una medida relativa, ya que, a diferencia del estimador del error estándar tradicional, elimina el efecto de la magnitud de la estimación. Lo anterior nos permite evaluar la calidad de las estimaciones y compararlas, a pesar de que hubieran sido medidas en escalas o unidades diferentes.

Coeficiente de variación, confiabilidad y error relativo

Sea X_1, X_2, \dots, X_n una muestra de variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 . Por el Teorema Central del Límite sabemos que la distribución de la media, \bar{X} , converge en distribución a una variable aleatoria normal, con media μ y varianza $\frac{\sigma^2}{n}$ (lo mismo aplicaría en el caso de la suma de las variables aleatorias, que convergería en distribución a una normal con media $n\mu$ y varianza $n\sigma^2$).

La convergencia nos permite construir intervalos de confianza simétricos centrados en el valor de la estimación. Por facilidad, utilizaremos $\mu \pm \frac{Z_{1-\alpha/2}\sigma}{\sqrt{n}}$, como una descripción de un intervalo de confianza, aunque debemos recordar que el intervalo contiene a todos los valores que se encuentran entre el límite inferior y el superior.

Podemos reescribir el intervalo de confianza como:

$$\mu \pm \frac{Z_{1-\alpha/2}\sigma}{\sqrt{n}} \equiv \mu \pm Z_{1-\alpha/2} \frac{\sigma\mu}{\mu\sqrt{n}} \equiv \mu \pm \mu Z_{1-\alpha/2} CV(\bar{X}) \equiv \mu \left(1 \pm Z_{1-\alpha/2} CV(\bar{X}) \right)$$

y en términos del CV, quedaría como sigue:

$$\mu \left(1 \pm Z_{1-\alpha/2} CV(\bar{X}) \right) \quad (1)$$

Por otro lado, un intervalo de confianza construido en términos del error relativo de la estimación, se puede expresar como:

$$\mu \pm r_{\alpha}\mu \equiv \mu(1 \pm r_{\alpha}) \quad (2)$$

De (1) y (2), obtenemos la siguiente igualdad:

$$\mu(1 \pm r_{\alpha}) = \mu \left(1 \pm Z_{1-\alpha/2} CV(\bar{X}) \right) \Rightarrow r_{\alpha} = Z_{1-\frac{\alpha}{2}} CV(\bar{X})$$

y

$$CV(\bar{X}) = r_{\alpha} / Z_{1-\frac{\alpha}{2}}$$

¿Pero cómo se podría interpretar lo anterior? Por ejemplo, si tenemos una estimación con un error relativo del 15% y una confiabilidad del 90%, (es decir, $r=.15$ y $\alpha = .1$), entonces el coeficiente de variación de la estimación es:

$$CV(\bar{X}) = r / Z_{1-\frac{\alpha}{2}} = \frac{.15}{1.644854} \approx 0.0911$$

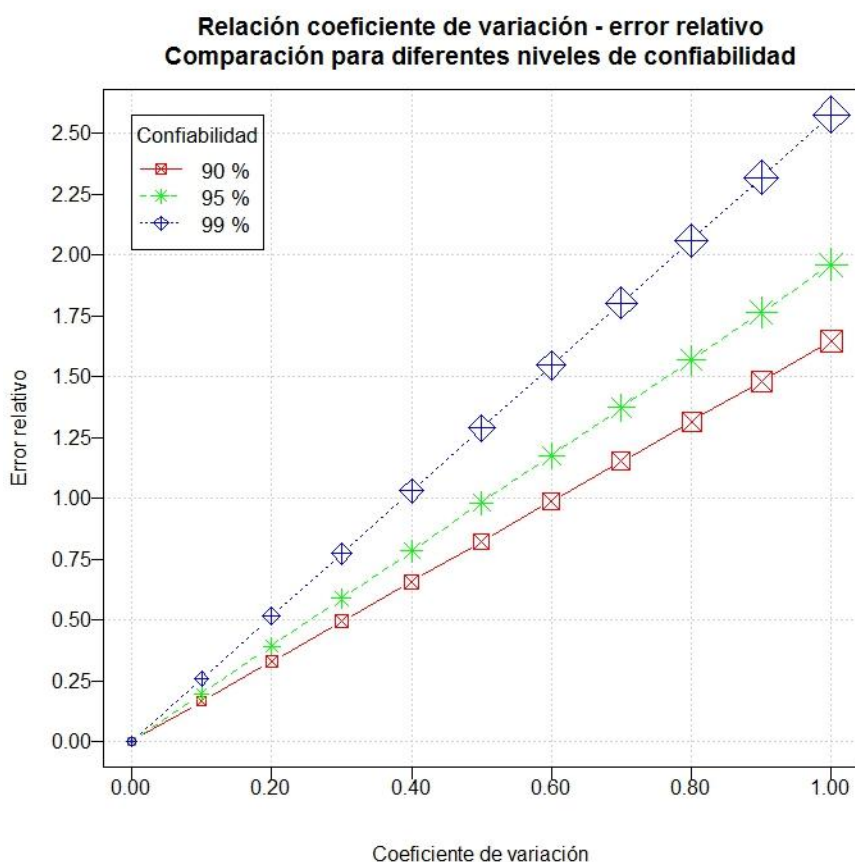
O aproximadamente del 9.1%.

Siguiendo la relación anterior, podemos construir una tabla que muestre la relación directa que existe entre el error relativo y el coeficiente de variación, bajo diferentes niveles de confiabilidad.

Coeficiente de variación	Error relativo para un nivel de confianza y coeficiente de variación dados		
	$1-\alpha = .90$	$1-\alpha = .95$	$1-\alpha = .99$
0.05	0.0822	0.0980	0.1288
0.10	0.1645	0.1960	0.2576
0.15	0.2467	0.2940	0.3864
0.20	0.3290	0.3920	0.5152
0.25	0.4112	0.4900	0.6440
0.30	0.4935	0.5880	0.7727
0.35	0.5757	0.6860	0.9015
0.40	0.6579	0.7840	1.0303
0.45	0.7402	0.8820	1.1591
0.50	0.8224	0.9800	1.2879
0.55	0.9047	1.0780	1.4167
0.60	0.9869	1.1760	1.5455
0.65	1.0692	1.2740	1.6743
0.70	1.1514	1.3720	1.8031
0.75	1.2336	1.4700	1.9319

Nota: Valores calculados en términos de proporciones.

La información de la tabla anterior se puede mostrar gráficamente, en donde se puede observar el efecto del nivel de confianza (mostrado en forma de proporciones):



A continuación, se muestran ejemplos de las interpretaciones de los resultados anteriores:

Ejemplo 1. Tamaño de muestra.

En la muestra de la ENIGH 2014, se obtuvo una estimación del promedio del ingreso corriente a nivel nacional de \$ 46,564. Si hubiéramos calculado el tamaño de muestra para estimarlo a nivel nacional, con un error relativo de 15% y una confiabilidad del 90%, entonces obtendríamos un coeficiente de variación para dicha estimación, de $CV(\bar{R}) = .15 / 1.644854 = 0.09119352$, o aproximadamente 9.1%.

El intervalo de confianza esperado, sería (39,579.40, 53,548.60).

Ejemplo 2. ¿Qué tan grande es un coeficiente de variación de 0.30 (30%)?

Retomando el ejemplo de la estimación del promedio del ingreso corriente en la ENIGH 2014 de \$ 46,564; supongamos que se obtuvo un coeficiente de variación de 0.30 (30%). Basándonos en un nivel de confianza de 90%, tenemos un error

relativo de .4935 (49.35%) lo que, traducido a un intervalo de confianza, se convierte en el siguiente rango: (23,584.66, 69,543.33).

Como se puede observar en este ejemplo, el intervalo de confianza asociado a la estimación es bastante amplio y, por tanto, impreciso. Recordar que todos los valores al interior del intervalo, son posibles verdaderos valores del estimador, bajo ese nivel de confianza.

Ejemplo 3. Comparación de coeficiente de variación en estimaciones consideradas como precisas.

Supongamos que se llevaron a cabo dos encuestas de forma paralela, las cuales miden el ingreso. En ambos casos se obtiene una estimación del ingreso similar de \$ 46,564; la primera tiene un coeficiente de variación de 10% y la segunda de 2%. Si los análisis se llevan a cabo con un nivel de confianza del 90% ¿cuál estimación será más confiable? Obviamente aquella con un coeficiente de variación menor, sin embargo, en términos del intervalo de confianza, se traduciría como sigue:

1. El error relativo de la primera estimación es 0.164 (16.4%), lo nos lleva a un intervalo de (38,927.50, 54,200.49).
2. El error relativo de la segunda estimación es de 0.033(3.3%), obteniendo un intervalo de (45,027.38, 48,100.61)

Ambas estimaciones serían consideradas como estimaciones aceptables dada la precisión del coeficiente de variación; sin embargo, se puede apreciar la gran diferencia entre los anchos de los intervalos de confianza.

Ejemplo 4. Coeficiente de variación de 0.25 (25%). Límite superior en la DGES.

Suponga que tenemos una estimación de una proporción de 0.40 asociado a coeficiente de variación de 0.25 (25%). Bajo un nivel de confianza de 90%, obtenemos un error relativo de 0.411 (41.1%). El intervalo de confianza asociado es de (0.2356, 0.5644). Debe ser claro que el ancho del intervalo es muy grande, e incluye valores tan dispares, que podrían cambiar la decisión sobre la política pública que esté asociada a ese dato.

Ejemplo 5. Estimaciones imprecisas 1.

A partir de una muestra se estimó que la proporción de personas que sufren violencia doméstica es del 0.03 (3%), sin embargo, la estimación tiene un coeficiente de variación de 45%. Suponiendo un nivel de confianza de 0.95 (95%), ¿cómo afecta eso a los intervalos de confianza?

En la tabla anterior, para un coeficiente de variación del 0.45 (45%) con un nivel de confiabilidad del 95%, se obtiene un error relativo de 0.882 (88.2%),

por tanto, el intervalo de confianza se puede escribir como (0.00354,.05646). Como se puede observar el intervalo garantiza al menos que la estimación es estadísticamente mayor que cero, sin embargo, no da mucha certeza del valor real de la proporción. En estos casos es necesario recordar el resto de indicadores de precisión publicados, para evaluar si nos sirve o no la estimación.

Ejemplo 6. Estimaciones imprecisas 2.

A partir de una muestra se estimó que la proporción de personas con cierta característica es del 0.53 (53%). Se obtuvo un coeficiente de variación de 0.45 (45%). Se pretende analizar la información con un nivel de confianza del 95%; ¿cómo afecta esto los intervalos de confianza?

Como podemos observar en la tabla un coeficiente de variación del 45% con un nivel de confiabilidad del 95% equivalen a un error relativo de .882 (88.2%), por tanto, el intervalo de confianza equivalente es (.06254%, .99746%), como se puede observar el intervalo garantiza al menos que la estimación es estadísticamente diferente de cero, sin embargo, la precisión del presente intervalo es muy cuestionable. Por esta razón, el Instituto no recomienda el uso de estimadores con valores del coeficiente de variación tan grandes.