

INCONSISTENCIA DEL COEFICIENTE DE VARIACIÓN PARA EXPRESAR LA VARIABILIDAD DE UN EXPERIMENTO EN UN MODELO DE ANÁLISIS DE VARIANZA

Inconsistency of the coefficient of variation for expressing the variability of an experiment in a model of analysis of variance

Edison Ramiro Vásquez[✉] y Alberto Caballero Núñez

ABSTRACT. From a variable distributed normally with mean zero and variance one, arose a succession of p treatments with constant variance equal to one and whose mean increased k units, valid to form two separate experiments; one with the t first treatments of the succession and the other with the remaining p ; It was found that despite the fact that both experiments maintained equals its estimate of experimental error variance, its variation coefficients are different. As a practical illustration were analyzed three experiments through an analysis of variance, which had the same variance and different structure of half of its four treatments. The three experiments presented different coefficients of variation in spite that their estimates of variance of the experimental error were the same, which shows that this statistician does not accurately reflect the variability existing in an experiment when it is parsed through an analysis of variance. Also the coefficient of variation was associated with the level of significance or capability of the F test to detect significant differences between treatments, or the range between the values of the averages of treatments.

Keywords: coefficient of variation, variation of the experimental error, analysis of variance

RESUMEN. A partir de una variable normalmente distribuida con media cero y varianza uno, se generó una sucesión de p tratamientos con varianza constante igual a uno y cuyas medias se incrementaban en k unidades, válidas para conformar sendos experimentos; uno con los t primeros tratamientos de la sucesión y otro con los $p-t$ restantes; se comprobó que a pesar que ambos experimentos mantenían iguales sus estimaciones de varianza de error experimental, sus coeficientes de variación son diferentes. Como ilustración práctica se analizaron tres experimentos mediante un Análisis de Varianza, que tenían la misma varianza y diferente estructura de medias de sus cuatro tratamientos. Los tres experimentos presentaron diferentes coeficientes de variación a pesar que sus estimaciones de varianza del error experimental eran las mismas, lo cual pone en evidencia que este estadígrafo no refleja exactamente la variabilidad existente en un experimento cuando es analizado a través de un Análisis de Varianza. Tampoco el coeficiente de variación se asoció con el nivel de significación o capacidad de la prueba de F para detectar diferencias significativas entre medias de tratamientos, ni con el rango entre los valores de las medias de tratamientos.

Palabras clave: coeficiente de variación, varianza del error experimental, análisis de varianza

INTRODUCCIÓN

Los estadígrafos son expresiones de la muestra, los cuales, si son concebidos e interpretados de manera correcta, pueden brindar información muy valiosa sobre la población de la cual fue extraída la muestra, como lo señalan (1, 2, 3, 4, 5).

El coeficiente de variación de Pearson, ha resultado ser una medida de dispersión de amplia utilización, cuando se pretende comparar la dispersión entre varias poblaciones,

Dr.C. Edison Ramiro Vásquez, Profesor Principal, Universidad Nacional de Loja (UNL), Ecuador y Dr.C. Alberto Caballero Núñez, Investigador Titular del departamento de Matemática Aplicada, Instituto Nacional de Ciencias Agrícolas (INCA), gaveta postal 1, San José de las Lajas, La Habana, Mayabeque, Cuba. CP 32 700.

✉ eramirov@hotmail.com

de una o diferentes variables medidas en la misma o diferentes escalas. Este estadígrafo es definido como la relación por cociente que se establece entre la desviación estándar y la media aritmética de la variable; dicho de otro modo, desviación estándar expresada como porcentaje de la media aritmética (6, 7, 8); esto lo hace un coeficiente adimensional, invariante por la escala de medición de la variable analizada. Todo lo anterior justifica la preferencia de utilizarlo cuando el propósito en la investigación sea la comparación de poblaciones desde el punto de vista de su variabilidad (9, 10, 11).

Sin embargo, el hecho de estar conformado por una razón entre dos estadígrafos que reflejan diferentes características en la población, cualquier interpretación acerca de este coeficiente debe ir unida a un análisis de las peculiaridades del problema tratado y el comportamiento

de cada uno de estos dos estadígrafos. En este contexto, es preciso tener en cuenta que la desviación estándar es invariante ante la relación suma, no así la media; de otro modo, si a cada valor de una muestra se le suma una constante k , la muestra resultante conserva su desviación estándar; y su media se afecta en k unidades; por tanto, su coeficiente de variación será diferente al de la muestra inicial.

Teniendo en cuenta lo anterior, se puede esperar que en los modelos de Análisis de Varianza, en los que se analizan sendos experimentos con similar estimación de varianza del error experimental, pero con diferente composición de medias de tratamientos, que pueden aportar medias generales distintas, los coeficientes de variación pueden resultar diferentes.

En este contexto, el objetivo del presente trabajo está dirigido a ilustrar la inconsistencia que tiene el Coeficiente de Variación para explicar la variabilidad de los experimentos analizados según un modelo de Análisis de Varianza.

MATERIALES Y MÉTODOS

Para cumplir con el objetivo propuesto se hace uso de la propiedad que plantea: dada una variable aleatoria X normalmente distribuida con media $\mu_x=0$ y varianza $\sigma_x^2=1$ si a cada uno de los valores que componen la población se le suma una constante $k>0$, se origina una nueva variable aleatoria que mantiene la varianza y la media se incrementa en la constante k .

Para el desarrollo teórico: Se generó una sucesión de p poblaciones de variables aleatorias $X_{(1)}, X_{(2)}, \dots, X_{(p)}$ de modo que $X_{(i)} = X_{(i-1)} + k$.

$$\text{con } \mu_{x(i)} = (i-1)k \text{ y } \sigma_{x(i)}^2 = 1 \quad i = 1, 2, \dots, p$$

Las p poblaciones $X_{(1)}, X_{(2)}, \dots, X_{(i)}, \dots, X_{(p-1)}, \dots, X_{(p)}$ se utilizaron para conformar los tratamientos de dos experimentos; Experimento 1 y Experimento 2, siguiendo la siguiente estructura:

Experimento 1. Sus tratamientos se corresponden con las t primeras poblaciones de la sucesión:

Medias de tratamientos: $0, k, 2k, \dots, (t-1)k$

$$\text{Media general: } \mu_{\text{Exp1}} = \frac{(t-1)k}{2}$$

$$\text{Varianza: } \sigma_{\text{Exp1}}^2 = 1$$

$$\text{Coeficiente de variación: } CV_{\text{Exp1}} = \frac{1}{\frac{(t-1)k}{2}} = \frac{2}{(t-1)k}$$

Experimento 2. Sus tratamientos se corresponden con las t últimas poblaciones de la sucesión:

Medias de tratamientos: $(p-t)k, (p-t+1)k, \dots, (p-1)k$

$$\text{Media general: } \mu_{\text{Exp2}} = kp - \frac{(t-1)k}{2} = \frac{k(2p-t-1)}{2}$$

$$\text{Varianza: } \sigma_{\text{Exp2}}^2 = 1$$

Coeficiente de variación:

$$CV_{\text{Exp2}} = \frac{1}{\frac{k(2p-t-1)}{2}} = \frac{2}{k(2p-t-1)}$$

Para la ilustración práctica. Se generó una población (A) normalmente distribuida con media 10 y varianza 25; a partir de esta, se generaron nueve poblaciones (B, C, ..., J), cuyas medias se incrementaron respectivamente en un valor de $k=2, 4, \dots, 18$ que conservan la misma varianza (Tabla I).

Tabla I. Conformación de poblaciones

k_i	2	4	6	8	10	12	14	16	18	
$x(i)$	A	B	C	D	E	F	G	H	I	J
1	8	10	12	14	16	18	20	22	24	26
2	10	12	14	16	18	20	22	24	26	28
3	9	11	13	15	17	19	21	23	25	27
4	8	10	12	14	16	18	20	22	24	26
5	3	5	7	9	11	13	15	17	19	21
6	6	8	10	12	14	16	18	20	22	24
7	7	9	11	13	15	17	19	21	23	25
8	15	17	19	21	23	25	27	29	31	33
9	20	22	24	26	28	30	32	34	36	38
10	14	16	18	20	22	24	26	28	30	32
Media	10	12	14	16	18	20	22	24	26	28
Varianza	25	25	25	25	25	25	25	25	25	25
CV (%)	50	42	36	31	28	25	23	21	19	18

Se conformaron tres experimentos (Tabla II)

Experimento 1. Tratamientos A, B, C, D (cuatro primeras poblaciones)

Experimento 2. Tratamientos G, H, I, J (cuatro últimas poblaciones)

Experimento 3. Tratamientos A, E, F, J (la primera, las dos del centro y la última población)

RESULTADOS Y DISCUSIÓN

Desarrollo teórico. Para comprobar la igualdad o no entre los coeficientes de variación se calculó el cociente entre ellos:

$$\frac{CV_{\text{Exp1}}}{CV_{\text{Exp2}}} = \frac{\frac{2}{k(t-1)}}{\frac{2}{k(2p-t-1)}} = \frac{k(2p-t-1)}{(t-1)k} = \frac{2p-t-1}{(t-1)}$$

En efecto, siendo $p>t$, hace que $CV_{\text{Exp1}} \neq CV_{\text{Exp2}}$

Este estadígrafo, definido como el cociente entre la desviación estándar de la muestra y su media, ofrece una estimación de la proporción de variabilidad de la muestra expresada por unidades de media y no la variabilidad en sí de dicha muestra (6, 7).

Tabla II. Estructura de los experimentos a partir de la Tabla I

Observaciones	Experimento 1				Experimento 2				Experimento 3			
	T ₁	T ₂	T ₃	T ₄	T ₁	T ₂	T ₃	T ₄	T ₁	T ₂	T ₃	T ₄
1	8	10	12	14	20	22	24	26	8	16	18	26
2	10	12	14	16	22	24	26	28	10	18	20	28
3	9	11	13	15	21	23	25	27	9	17	19	27
4	8	10	12	14	20	22	24	26	8	16	18	26
5	3	5	7	9	15	17	19	21	3	11	13	21
6	6	8	10	12	18	20	22	24	6	14	16	24
7	7	9	11	13	19	21	23	25	7	15	17	25
8	15	17	19	21	27	29	31	33	15	23	25	33
9	20	22	24	26	32	34	36	38	20	28	30	38
10	14	16	18	20	26	28	30	32	14	22	24	32
Media	10	12	14	16	22	24	26	28	10	18	20	28
Varianza	25	25	25	25	25	25	25	25	25	25	25	25
CV (%)	50	42	36	31	23	21	19	18	50	28	25	18

T₁, T₂, T₃, T₄: tratamientos

De aquí que en variables previamente estandarizadas, este estadígrafo no puede ser definido (8); y en variables cuya media se aproxime a cero, su valor tiende a elevarse.

Los dos experimentos poseen igual varianza del error experimental, lo cual significa que la variabilidad entre unidades experimentales dentro de los respectivos tratamientos en cada experimento es la misma; no obstante, como sus medias generales son diferentes, el Experimento 1 posee mayor coeficiente de variación que el Experimento 2.

Lo anterior está asociado con la propia definición de este estadígrafo, que además de ser una función de la varianza, lo es también de la media, característica que lo hace variable ante los cambios de origen de escala, como lo señalan muchos autores, entre otros (6, 7).

Ilustración práctica. En la Tabla III se exponen los resultados del Análisis de Varianza de un factor con cuatro tratamientos y 10 repeticiones.

Tabla III. Resultados del ANOVA de los tres experimentos

Experimento	Media general	CME	P-ANOVA	CV (%)	Rango de medias
1	13	24,89	0,062	38	6
2	25	24,89	0,062	20	6
3	19	24,89	0,000	26	18

Como se aprecia, en tres experimentos los cuadrados medios de error; estimación de la varianza del error experimental o la varianza entre unidades experimentales dentro de cada tratamiento, son iguales; mientras que sus respectivas medias generales son diferentes, ello hace que las estimaciones de Coeficientes de Variación también lo sean, mayor en el experimento con menor

media general. Esto se debe a que la media general está notablemente influenciada por la estructura de las *t* medias de tratamientos que conforma el experimento (Tabla II).

Cabe resaltar que en el experimento analizado a través de un modelo de Análisis de Varianza el coeficiente de variación, expresa la proporción de la variabilidad del error experimental por unidades de media general y no el de la varianza del error experimental o la variabilidad absoluta del error del experimento. Lo cual evidencia que este estadígrafo es incapaz de reflejar en sí la variabilidad de dicho experimento, de aquí el cuidado que se debe tener en su interpretación ante estas situaciones.

Dado que los tres experimentos presentan igualdad en las estimaciones de varianza del error experimental, existe una clara asociación entre el nivel de significación con el cual se declaran diferencias entre medias de tratamientos a través del Análisis de Varianza y el rango entre dichas medias, a favor del experimento que posee rango mayor.

No se observa asociación entre el Coeficiente de Variación ni la varianza del error experimental con los rasgos descritos anteriormente; por tanto, estos estadígrafos no ofrecen una idea exacta de la capacidad del ANOVA para detectar una diferencia significativa entre medias de tratamientos; tampoco el rango de variación entre las medias de tratamientos constituye un indicador que refleje la variabilidad del error experimental.

Todo lo anterior pone en evidencia la poca sostenibilidad de establecer rangos en los que se debe ubicar este estadígrafo en un experimento para decidir acerca la calidad del experimento en su ejecución.

Resulta imprescindible hacer ciertas precisiones relacionadas con los resultados de este trabajo y el uso que se ha hecho del Coeficiente de Variación.

Es indudable que la elección de este estadígrafo es obligada cuando se pretende comparar la variabilidad de dos o más poblaciones medidas en igual o diferente escala; ningún otro estadígrafo de variabilidad es capaz de ofrecer información tan valiosa en tales situaciones.

Enumerar los trabajos en los que se hace uso acertado de este estadígrafo y su correcta interpretación, puede resultar un imposible. En tal sentido, solo se expondrán algunas situaciones en el que el uso del coeficiente de variación ha logrado extraer valiosa información para la población en disímiles campos de investigación: la Oficina Nacional de Estadística de Cuba ha evaluado la precisión de los resultados en cuanto a la calidad y la cobertura del censo de población y vivienda del 2002 en Cuba (9) a través de un uso eficiente de este estadígrafo; del mismo modo el Departamento Administrativo Nacional de Estadística, DANE, de Colombia (10) hace uso de este estadígrafo para medir el error muestral durante el censo de población del 2005. Otro ejemplo no menos importante se presenta (11) con la utilización de este estadígrafo para evaluar el nivel de uniformidad en la distribución de los abonos.

De todo lo anterior se recomienda que en modelos de Análisis de Varianza, no se emplee el Coeficiente de Variación como indicador de la variabilidad experimental, ni establecer rangos de sus magnitudes para asociarlas con la calidad de conducción del experimento.

REFERENCIAS

1. Little, T. y Jackson, F. 1991. Métodos estadísticos para la investigación en la agricultura. 2 ed. México: Editorial Trillas. 270 p.
2. Johnson, R. 1997. Probabilidad y Estadística para Ingenieros. 5 ed. México. : Prentice Hall.
3. Mendenhall, W. y Sincich, T. 2007. Statistics for engineering and the sciences. 5 ed. México: Prentice Hall. 1060 p.
4. Peña, D. 2008. Fundamentos de estadística. Madrid: Alianza Editorial. 688 p.
5. Vilar, J. 2011. Curso de Estadística 2 La Coruña: Universidad de La Coruña. [Consultado: 15 febrero 2011]. Disponible en: <http://www.udc.es/dep/mate/estadistica2/estadistica_2.htm>.
6. Berón López, F. J. Bioestadística: Métodos y Aplicaciones U.D. Bioestadística. Facultad de Medicina. Universidad de Málaga. ISBN: 847496-653-1.
7. Cabrera, G. F. Propiedades del coeficiente de variación (www.monografias.com)
8. Ruiz y Sánchez. Apuntes de Estadística. Edición electrónica. [Consultado: 01-2011]. Disponible en: www.eumed.net/libros/2006/rmsf/. 2006.
9. Oficina Nacional de Estadística, Cuba. Censo de Población y Vivienda. Aplicaciones del muestreo en el Censo de Población y Viviendas. [Consultado: 01-2011]. Disponible en: <http://www.cubagob.cu/otras_info/censo/c_vi.htm>. 2002
10. Estimación e interpretación del Coeficiente de Variación de la encuesta cocensal. Censo General 2005-CGRAL. Junio 2008. [Consultado: 01-2011]. Disponible en: <http://www.dane.gov.co/files/investigaciones/boletines/censo/est_interp_coefvariacion.pdf>.
11. Gil, E. Elementos clave en la uniformidad de distribución de abonos. Escuela Superior de Agricultura de Barcelona. [Consultado: 01-2011]. Disponible en: <<http://www.e-md.upc.edu/diposit/material/22459/22459.pdf>>.

Recibido: 18 de marzo de 2011

Aceptado: 8 de junio de 2011