COEFICIENTE DE VARIACIÓN Y PUNTAJES TÍPICOS
(Guía de clase)

FICHA Nº 19

A. Esta ficha de actividad tiene como objetivo presentar dos pequeños temas vinculados al uso conjunto de las medidas de tendencia y de dispersión que dan lugar a transformaciones y estadísticos de gran utilidad en el análisis estadístico.

- El origen de ambas se encuentra en la necesidad de comparar distribuciones univariadas para dos variables distintas o para la misma variable pero relevada en dos momentos distintos.
- Es sabido que tanto la comparación de promedios como la comparación de las dispersiones dependerá de las unidades en que ha sido medida, de la forma de la distribución que tienen las variables en el análisis, del promedio y de la varianza. Por lo tanto, habrá que diseñar formas algebraicas que permitan controlar estos aspectos.
- Cuando el interés se concentra en responder a la pregunta por cuál variable muestra una distribución más o menos homogénea, el uso del coeficiente de variación puede ser de gran utilidad. Si el propósito es contar con indicadores comparables que no estén afectados por las unidades de medida o métricas, o si han sido relevados en distintos momentos, la construcción de puntuaciones típicas será apropiado. Se verá que cada alternativa tiene sus ventajas y limitaciones.

B. Si el objetivo del análisis es comparar la heterogeneidad en dos distribuciones para una misma variable X pero extraídas de dos poblaciones o momentos diferentes en el tiempo, es necesario contar con medidas de dispersión relativas.

COEFICIENTE DE VARIACIÓN Y PUNTAJES TÍPICOS
(Guía de clase)

✎ Todas las medidas de dispersión presentadas en la Ficha nº18 pueden ser tratadas como “absolutas” dado que están atadas al valor concreto que tome la media aritmética en una distribución.

i) Es de recordar que tanto la varianza como el desvío estándar se calculan tomando las desviaciones a la media aritmética de la distribución.

ii) Si la pregunta es por cuál de las dos poblaciones o grupos es más heterogénea en X, la comparación no podrá hacerse directamente a menos que las medias aritméticas de ambas distribuciones sean exactamente iguales.

iii) Para la comparación de las dispersiones en el caso de medias aritméticas distintas, será necesario controlar por el valor de la media.

✎ El **coeficiente de variación** es un estadístico que combina una medida de dispersión (el desvío estándar) con una medida de tendencia (la media) para una misma distribución.

- Formalmente:

$$CV = \frac{S}{\overline{X}}$$

- Obsérvese que se emplea el desvío y no la varianza para su cálculo.

- En el cuadro 19.1 se muestran estadísticos distintos para las escuelas primarias de Argentina, México y Uruguay, para los sectores institucionales público urbano, público rural y privado. Si se observa, el desvío para las escuelas públicas urbanas argentinas es mayor que para el sector rural de ese país; sin embargo, al controlar por las medias respectivas, la imagen que se forma es exactamente la contraria: hay mayor heterogeneidad en el sector rural que en el urbano.

COEFICIENTE DE VARIACIÓN Y PUNTAJES TÍPICOS
(Guía de clase)

Cuadro 19.1
Tamaño de las escuelas primarias en número de alumnos

	Argentina				**México**				**Uruguay**			
	Pub. Urbana	**Rural**	**Privada**	**total**	**Pub. Urbana**	**Rural**	**Privada**	**total**	**Pub. Urbana**	**Rural**	**Privada**	**total**
Promedio de alumnos por escuela	487,0	152,1	357.2	372,5	408,3	110,9	310,1	235,9	456,3	61,0	381,0	388,1
Mínimo	50	29	31	29	8	1	11	1	94	14	104	14
Máximo	910	850	932	932	1437	856	1248	1437	1012	142	1155	1155
Devío estándar	202,5	139,2	193,4	237,0	236,0	119,0	293,3	234,5	158,2	40,2	325,1	231,6
Coeficiente de variación (%)	41,6	91,5	54.1	63,6	57,8	107,3	94,6	99,4	34,7	65,9	85,3	59,7

FUENTE: elaboración propia sobre la base de los microdatos de ONE (1999); TIMSS 99-R (1998); EN 4to. (2001); y UMRE (1999) respectivamente.

✎ Resulta importante señalar que el coeficiente de variación es un estadístico: una medida que resume una característica de una distribución.

- Al igual que la media o que la varianza puede ser calculado para toda la población analizada o para subgrupos de la población. Por ejemplo, valores de Y dentro de las categorías de la variable X.

- Este tipo de comparación será una comparación que no permite trabajar directamente con los valores que la variable toma en las unidades de registro.

- Si se desea realizar una comparación que transforme los valores de la o de las variables de interés, será necesario pasar a puntuaciones tipicas.

C. Los puntajes típicos también denominados puntajes z, constituyen una transformación algebraica de los valores de una variable de interés para expresarlos en una nueva variable que tendrá una media 0 y un desvío estandar de 1.

✎ Su utilidad consiste en que permite comparar valores de una o varias variables suprimiendo el efecto que de las distintas métricas:

- Un ejemplo que suele utilizarse es la comparación de los puntajes obtenidos por los alumnos en dos tipos de pruebas de aprendizaje (por ejemplo, matemática y ciencias) pero que se caracterizan por tener distinta extensión (en número de ítemes o ejercicios). Es claro que si la prueba de matemática tiene 35 ejercicios y la prueba de ciencias 45, no se podrá responder directamente sobre cuál es el área de conocimientos en la que los alumnos muestran más alta competencia. Los promedios respectivos pueden diferir porque la extensión de la prueba es distinta y no porque hayan diferencias reales. Al transformar las variables a puntajes típicos estas diferencias desaparecen.

- Otro ejemplo podría consistir en la comparación de municipios respecto a un conjunto de indicadores sociales. Si se deseara conocer qué municipios están por encima de la media y qué tan apartados están de la media en cada indicador, sería necesario realizar procesamientos específicos para cada variable. Pero, si se transforman las variables a puntajes z, se podrá responder directamente a estas preguntas en forma muy sencilla.

- Conceptualmente, no se trata de un estadístico, sino de una transformación de la variable original.

✎ Un puntaje típico se computa aplicando la siguiente fórmula a cada una de las X variables que se interesa transformar:

COEFICIENTE DE VARIACIÓN Y PUNTAJES TÍPICOS
(Guía de clase)

$$z_i = \frac{x_i - \bar{x}}{s}$$

- En la expresión anterior:

  I. X_i representa el valor de la variable en la *i-ésima* unidad
  II. $\bar{x}$ representa la media aritmética de X
  III. s representa el desvío estándar de X
  IV. z_i representa el valor de la nueva variable z en la *i-ésima* unidad

- Los nuevos puntajes z se caracterizan por algunas propiedades:

  I. Cada nuevo valor estará expresado en DESVIACIONES A LA MEDIA como indica el denominador.

  II. El promedio de la nueva variable será igual a 0 por aplicación de la primera propiedad de la media aritmética (ver ficha 17).

  III. La nueva variable será por tanto, de tipo interval, ya que el valor 0 es un valor arbitrariamente fijado (en la media aritmética de la variable originaria).

  IV. El desvío estándar será igual a 1.

✎ La variable z es una transformación lineal de la variable originaria. En consecuencia, no se alterará la forma de la distribución original. Esto es una propiedad particularmente importante dado que si la distribución original no era normal su transformación en puntaje z no le otorgará dicha forma.

COEFICIENTE DE VARIACIÓN Y PUNTAJES TÍPICOS
(Guía de clase)

✎ Corrientemente se afirma que los puntajes z constituyen una estandarización de los valores de una variable en la medida en que se controlan por el desvío estándar computado en la población bajo análisis. Esto tiene una consecuencia muy importante para la comparación que ha sido sistemáticamente desconocida en varios campos de la evaluación de las políticas sociales, muy en particular en la educación.

- El caso es que una muy frecuente estrategia de análisis de los niveles de aprendizaje en los alumnos evaluados ha sido estandarizar los resultados. Tal es el caso por ejemplo, de Chile con el Sistema de Medición de la Calidad de la Educación (SIMCE) y de la Dirección General de Educación (DGE) de México, hasta el 2003. Como la estandarización se hace utilizando parámetros de una población / año evaluado, no es posible responder a lo largo de los años, si los alumnos han mejorado o empeorado su nivel de aprendizaje. El 0 para cada año será igual a la media observada en ese año, sin posibilidades de enunciar si es diferente a la observada en años anteriores. Es decir, no se pueden realizar comparaciones interanuales, cuestión que ha sido uno de los usos más importantes a los que se ha querido aplicar todo sistema de evaluación de aprendizajes. Para este caso particular, resultan apropiados los valores originales (conocidos en la evaluación como puntajes brutos).