






MEDIDAS DE DISPERSIÓN (Guía de clase)

A. Una pregunta muy particular que se puede hacer a una distribución de datos es de qué magnitud es la heterogeneidad que se observa.

FICHA N° 18

-  Las medidas de dispersión generalmente acompañan a las medidas de tendencia central por una razón básica: la media aritmética informa sobre cuál es el centro de gravedad pero qué tan homogénea es la distribución.
-  Conceptualmente, la pregunta por una estimación de la heterogeneidad está en la base de los estudios de desigualdad en la distribución del ingreso, de bienes culturales, del acceso a servicios sociales.
-  La comparación de las medidas de desigualdad entre dos colectivos (por ejemplo, desigualdad de aprendizajes entre escuelas; ingresos entre países; de ingresos entre años distintos) constituye un enfoque problema de investigación que se pregunta en qué difieren dos casos de interés.
-  Por lo general, las medidas de dispersión se utilizan conjuntamente con las medidas de tendencia central. De hecho, en algunas de las fórmulas para su cálculo se emplea directamente la media aritmética como referencia de dispersión.
-  La diferencia entre las medidas obedece a una pregunta: ¿mayor o

MEDIDAS DE DISPERSIÓN (Guía de clase)

menor dispersión respecto de qué?.

- ❖ Las medidas que se han ido desarrollando constituyen sofisticaciones matemáticas a las sucesivas respuestas dadas a esta pregunta.
- ❖ Por principio, las medidas de dispersión ubican su referente de dispersión en la propia distribución observada de los datos y no en relación de un parámetro ideal o deseable. Este no es el caso en todas las medidas de desigualdad. En el caso del índice de Gini para medir concentración y desigualdad, su referente es ideal: una distribución completamente democrática.

B. Las medidas de dispersión más frecuentemente utilizadas son el rango, la varianza y el desvío estándar.





Una medida simple que se podría proponer para analizar la dispersión es el **rango**. Éste se define la distancia entre el valor mínimo y el máximo observados.


- ❖ Esta medida permitiría observar qué tan amplia es la distribución. Una medida importante por ejemplo, cuando se comparan tamaños de hogares en distintos países o regiones de América Latina.
- ❖ Tiene un inconveniente importante: el rango puede tomar valores máximos y mínimos que son **atípicos** (con una frecuencia simple de 1, por ejemplo) o muy desviados o “**out-layers**” (valores extremadamente altos o extremadamente bajos en comparación del resto de los valores).
- ❖ Otro inconveniente es que dos distribuciones pueden tener el mismo rango pero si se grafica la distribución, se observarían diferencias muy importantes.

MEDIDAS DE DISPERSIÓN


(Guía de clase)

-  La **varianza** es un estadístico que informa sobre cuál es la “distancia promedio” de las observaciones a la media aritmética de una distribución.
-  El **desvío estándar** se define como la raíz cuadrada de la varianza y suele ser más utilizada para fines descriptivos por razones de practicidad se utiliza más frecuentemente. En el campo de la estadística inferencial el desvío estándar presenta dificultades por lo que por lo general se usa en su la varianza.

C. La varianza y el desvío estándar se calculan siguiendo la misma idea de encontrar un promedio en las dispersiones de los valores observados respecto a su media.

-  La **varianza** se define como el promedio de los cuadrados de las desviaciones a la media aritmética y se expresa algebraicamente en la siguiente forma:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

-  El **desvío o desviación estándar** es la raíz cuadrada de la varianza. Como se apreciará en el cuadro 18.1, sus valores son sensiblemente más reducidos y por tanto manipulables.

MEDIDAS DE DISPERSIÓN (Guía de clase)

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

D. Algunas características de la varianza y del desvío son:



Los símbolos utilizados para identificar a la varianza pueden ser distintos según el contexto estadístico en que se los está utilizando:

- i) La letra griega “ σ^2 ” (sigma) se utiliza para indicar que la varianza ha sido calculada en una población (registrada en un censo).
- ii) Las letras latinas “ S^2 ” mayúscula y “ s^2 ” minúscula se emplean por lo general para referirse a las varianzas y también los desvíos pero sin el cuadrado) que se calculan en una muestra. En tal caso, se dirá que la varianza calculada en una muestra (“ s^2 ”) es un estimador de la varianza poblacional (“ σ^2 ”).
- iii) También puede representarse con la letra “V”.



La varianza nunca tendrá valores negativos.

- Esto se deriva de la propia fórmula, donde cualquiera sea el signo de las desviaciones a la media, al elevarse al cuadrado tomarán valores positivos (esta es la primera propiedad de la varianza). Formalmente:

$$S^2 \geq 0$$

MEDIDAS DE DISPERSIÓN (Guía de clase)

- Si la varianza toma valor 0, entonces se trata de una variable que toma el mismo valor en todas las unidades. Es decir, se trata de una **constante**. Si se recuerda que una de las propiedades de la media aritmética informa que la media de la constante es igual a la constante, formalmente se tiene que:

$$S^2(k) = \frac{\sum (k_i - \bar{k})^2}{N} = \frac{0}{N} = 0$$

Es de observarse que la anterior es la segunda propiedad de la varianza.



La varianza es sensible a la presencia de valores extremos, tal como lo es la media aritmética, aunque es más estable que esta. En el cuadro siguiente se ha presentado el mismo ejemplo de la Ficha nº17, y se ha calculado la varianza para cada una de las tres regiones.

E. Siguiendo a Cortés (2000:159 y ss) distinguiremos las siguientes otras cuatro propiedades de la varianza.



Tercera propiedad: al sumar una constante a los valores originales de la variable, la varianza no se modifica.

- Formalmente:

$$S^2(K + X) = S^2(X)$$

- En el cuadro anterior, la columna que identifica a la región “D” presenta una distribución del ingreso con los mismos valores que en la región “A” pero se le han agregado 100U\$S a cada hogar. Si se observa la

MEDIDAS DE DISPERSIÓN (Guía de clase)

media se ha incrementado en 100 U\$S (tercera propiedad de la media, Ficha nº17) pero la varianza y el desvío no se han modificado.

iii) Se demuestra que:

$$S^2(K + X) = \frac{\sum (K + X_i - K + \bar{X})^2}{N}$$

Reordenando:

$$S^2(K + X) = \frac{\sum (X_i - \bar{X} + K - K)^2}{N}$$

$$S^2(K + X) = \frac{\sum (X_i - \bar{X})^2}{N}$$

$$S^2(K + X) = S^2(X)$$



Cuarta propiedad: si se multiplican los valores de la variable por una constante, la varianza de la nueva variable resultante será igual a la varianza original por la constante.

i) Formalmente:

$$S^2(K * X) = K^2 * S^2(X)$$

ii) Por las propiedades de la media, se recuerda que en este caso la media de la nueva variable será igual a la media original por la constante. De aquí que:

MEDIDAS DE DISPERSIÓN (Guía de clase)

$$S^2(K * X) = \frac{\sum (K * X_i - K * \bar{X})^2}{N}$$

Reordenando:

$$S^2(K * X) = \frac{\sum [K(X_i - \bar{X})]^2}{N}$$

$$S^2(K * X) = \frac{K^2 \sum (X_i - \bar{X})^2}{N}$$

$$S^2(K * X) = K^2 * S^2(X)$$


MEDIDAS DE DISPERSIÓN
(Guía de clase)

Cuadro 18.1
Simulación de ingresos de los hogares en tres regiones
(Valores en dólares)

	Ingreso en la Región "A"	Ingreso en la Región "B"	Ingreso en la Región "C"	Ingreso en la Región "D"	Ingreso en la Región "E"
Hogar 1	828,2	828,2	828,2	928.2	8282.0
Hogar 2	675,0	675,0	675,0	775.0	6750.0
Hogar 3	603,3	603,3	603,3	703.3	6033.3
Hogar 4	804,8	804,8	804,8	904.8	8047.5
Hogar 5	400,0	400,0	400,0	500.0	4000.0
Hogar 6	1066,7	1066,7	1066,7	1166.7	10666.7
Hogar 7	2145,0	2145,0	2145,0	2245.0	21450.0
Hogar 8	635,5	635,5	635,5	735.5	6355.0
Hogar 9	760,5	760,5	760,5	860.5	7605.0
Hogar 10	525,0	525,0	525,0	625.0	5250.0
Hogar 11	1731,2	1731.2	1731,2	1831.2	17311.7
Hogar 12	279,0	279.0	279,0	379.0	2790.0
Hogar 13	275,0	275.0	275,0	375.0	2750.0
Hogar 14	482,6	482.6	483,0	582.6	4825.8
Hogar 15	475,2	475.2	475,2	575.2	4751.7
Hogar 16	333,3	333,3	333,3	433.3	3333.3
Hogar 17	525,0	525,0	525,0	625.0	5250.0
Hogar 18	781,8	781,8	781,8	881.8	7818.3
Hogar 19	992,3	992,3	992,3	1092.3	9922.5
Hogar 20	722,7	722,7	722,7	822.7	7226.7
Hogar 21	625,0	6250,0	62,5	725.0	6250.0
Hogar 22	1274,3	12743,0	127,4	1374.3	12743.3
Hogar 23	1373,5	13735,0	137,4	1473.5	13735.0
Media aritmética	796,29	2076,96	668,25	896.29	7962.95
Varianza	215490.70	13922922,52	236022,75	215490.70	21549069.56
Desvío estándar	464.21	3731.34	485.82	464.21	4642.10

Fuente: los valores para la región "A" han sido tomados de la Encuesta Continua de Hogares (ECH) del Uruguay para el año 2001 y corresponden a 23 hogares tomados de un departamento del Interior.

MEDIDAS DE DISPERSIÓN (Guía de clase)

 **Quinta propiedad:** si los valores originales de una variable se multiplican por una constante (K_1) y luego se le suma otra constante (K_2), la varianza de la nueva variable será igual al cuadrado de la constante K_1 por la varianza original.

i) Se demuestra que es una combinación de las dos anteriores

$$S^2(K_1 + K_2 * X) = \frac{\sum [(K_1 + K_2 * X) - (K_1 + K_2 * \bar{X})]^2}{N}$$

$$S^2(K_1 + K_2 * X) = \frac{\sum [(K_1 - K_1) + (K_2 * X) - (K_2 * \bar{X})]^2}{N}$$


$$S^2(K_1 + K_2 * X) = \frac{\sum [(K_2 * X) - (K_2 * \bar{X})]^2}{N}$$

$$S^2(K_1 + K_2 * X) = \frac{\sum [K_2 (X_i - \bar{X})]^2}{N}$$

$$S^2(K_1 + K_2 * X) = \frac{K^2 \sum (X_i - \bar{X})^2}{N}$$

$$S^2(K_1 + K_2 * X) = K^2 * S^2(X)$$

propiedades:

 **Sexta propiedad:** la varianza de un conjunto de observaciones clasificadas según las categorías de otra variable (nominal u ordinal) es igual a la suma de la intervarianza y la intravarianza.

i) La propiedad se conoce también con el nombre de “descomposición de la varianza” y es muy recurrida en distintas aplicaciones estadísticas.

MEDIDAS DE DISPERSIÓN (Guía de clase)

- ii) A pesar de que en su demostración están involucradas dos variables (la variable de interés, X, y la variable categórica, Z), en este caso puntual **no se trata de una forma de análisis bivariado**, aunque luego dará lugar a una técnica específica: el análisis de varianza.
- iii) En el cuadro 18.2 se presenta esquemáticamente la propiedad.
- iv) La varianza de X en cada categoría de Z, (S^2_j), se denomina varianzas internas; el promedio ponderado de éstas genera la **intravarianza** (S^2_w). La varianza que se calcula a partir de las medias aritméticas de cada una de las categorías de Z, se denomina **intervarianza** (S^2_B).

Cuadro 18.2
Esquema de la descomposición de la varianza

Variable Z	Número de casos de X en cada categoría de Z (n_i)	Promedios para X en cada categoría de Z	Varianzas de X en cada categoría de Z
Categoría A	N_a	\bar{x}_a	S^2_a
Categoría B	N_b	\bar{x}_b	S^2_b
Categoría C	N_c	\bar{x}_c	S^2_c
...
Categoría J	N_j	\bar{x}_j	S^2_j
Total	N	\bar{x}	S^2

MEDIDAS DE DISPERSIÓN

(Guía de clase)

iv) Formalmente, la propiedad establece que:

$$S^2_T = S_B^2 + S_W^2$$

v) Donde:

$$S^2_W = \frac{\sum S_i^2 * N_i}{N}$$

$$S^2_B = \frac{\sum (\bar{Y}_i - \bar{Y})^2}{N}$$