Aula360
Aprendizaje para el siglo XXI

Apuntes Unidad 2

Propiedades de las medidas de tendencia central y de dispersión

fcfm
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

CMM
Centro de Modelamiento Matemático

MEDIA Y DESVIACIÓN ESTÁNDAR DE DATOS DESPLAZADOS

Si a todos los datos de una muestra se les suma una misma cantidad a, la media muestral de los datos desplazados es la media de los datos originales más a. Análogamente, si a todos los datos se le resta una misma cantidad a, la nueva media se obtiene como la media de los datos originales menos a.

Es decir, si tenemos un conjunto de n datos $x_1, x_2, \ldots, x_n$, cuya media es $\bar{x}$, y a cada dato le sumamos una cantidad a, que puede ser positiva o negativa, se obtienen los valores $x_1 + a, x_2 + a, \ldots, x_n + a$ y la media de estos nuevos datos es:

$$\bar{x}_a = \bar{x} + a$$

Por otro lado, la desviación estándar de estos nuevos datos es la misma que la de los datos originales. Es decir, para los datos $x_1, x_2, \ldots, x_n$, con media igual a $\bar{x}$, se tiene que como la desviación estándar se calcula a partir de las diferencias de los datos y la media, y al desplazar los datos, estas diferencias no cambian, se obtiene la misma desviación estándar.

Una vez estudiado esto, podemos pensar en otra situación ficticia. Supongamos que a comienzo de año, estas tres personas ahorraron 5.000, 7.500 y 10.000 pesos en sus cuentas bancarias. A final de año, estas personas disponían del doble de lo inicialmente depositado, es decir, tenían 10.000, 15.000 y 20.000 pesos respectivamente. ¿Cómo cambia la media entre estos tres ahorros?¿Cambia la desviación estándar? Revisemos esto de manera más general.

MEDIA Y DESVIACIÓN ESTÁNDAR DE DATOS MULTIPLICADOS POR UNA MISMA CANTIDAD

Si todos los datos de una muestra se multiplican por una misma cantidad a, la media muestral de estos nuevos datos es la media de los datos originales multiplicada por a.

Es decir, si tenemos los datos: $x_1, x_2, \ldots, x_n$, con media $\bar{x}$, y se multiplican todos los datos por una cantidad a, se obtienen los valores: $a \cdot x_1, a \cdot x_2, \ldots, a \cdot x_n$ y la media de estos nuevos datos es:

$$\bar{x}_a = a \cdot \bar{x}$$

Por otro lado, la desviación estándar de estos nuevos datos es la de los datos originales multiplicada por a. Es decir, considerando los datos: $x_1, x_2, \ldots, x_n$ y su media $\bar{x}$, como la desviación estándar se calcula a partir de las diferencias de los datos respecto a la media al multiplicar todos los datos por a, la diferencias de estos nuevos datos respecto a su media $\bar{x}_a$ corresponden a las diferencias de los datos originales multiplicadas por a. A partir de esto se obtiene que $s_a = a \cdot s$.

En el caso que la cantidad a sea un valor negativo, al calcular la raíz de a^2 se obtiene que la desviación estándar de los datos multiplicados, s_a, es igual a la desviación estándar de los datos originales, s, multiplicada por el valor absoluto de a.

Es importante notar que estas propiedades no solo son válidas cuando los datos provienen de una muestra, sino que también son ciertas para cualquier conjunto de datos.

COEFICIENTE DE VARIACIÓN

El **Coeficiente de Variación** es una medida de dispersión que se calcula dividiendo la desviación estándar de la muestra por su media aritmética. Se suele expresar en porcentaje, por lo que se calcula de la siguiente forma:

$$CV = \frac{s}{|\bar{x}|} * 100\%$$

Notemos que el coeficiente de variación representa la desviación estándar con respecto a la media y que, como ambas medidas tienen la misma unidad, es un valor libre de unidades. Debido a esto, permite comparar la variabilidad de distintos conjuntos de datos, independiente de la magnitud y de la unidad de los datos. Mientras mayor sea la variabilidad de una muestra, mayor será su coeficiente de variación.

Al igual que lo visto en la lección anterior, veremos cómo se comporta el coeficiente de variación cuando los datos son multiplicados por una misma cantidad.

COEFICIENTE DE VARIACIÓN Y DATOS MULTIPLICADOS POR UNA MISMA CANTIDAD

Supongamos que los datos de una muestra A, expresados en una unidad, se multiplica por una constante a para obtener los datos expresados en la otra unidad, llegando a una muestra B. Como hemos visto, esta multiplicación se traduce en las siguientes relaciones:

$$\bar{x}_A = a * \bar{x}_B \qquad s_A = |a| * s_B$$

Cuando calculamos el coeficiente de variación para la muestra A, y ocupamos estas

$$CV_A = \frac{s_A}{|\bar{x}_A|} = \frac{|a| * s_B}{|a| * |\bar{x}_B|} = \frac{s_B}{|\bar{x}_B|} = CV_B$$

igualdades, obtenemos:

Por lo tanto, el coeficiente de variación de una muestra tendrá el mismo valor si los datos están expresados en unidades distintas que son proporcionales entre sí, es decir, el coeficiente de variación de variación permanece inmutable al multiplicar por una constante a.

Ahora, ¿Qué sucede con el coeficiente de variación si desplazamos los datos? Veamos a continuación.

COEFICIENTE DE VARIACIÓN Y DATOS DESPLAZADOS EN UNA MISMA CANTIDAD

Supongamos que los datos de una muestra A se desplazan en un valor a, obteniendo una muestra B. Según lo visto y utilizado previamente, se obtiene lo siguiente:

$$\bar{x}_A = \bar{x}_B + a \qquad\qquad s_A = s_B$$

Luego, usando estas relaciones, y asumiendo que a es distinto de 0, podemos ver que se tiene:

$$CV_A = \frac{s_A}{|\bar{x}_A|} = \frac{s_B}{|\bar{x}_B + a|} \neq \frac{s_B}{|\bar{x}_B|} = CV_B$$

Esto significa que, cuando tenemos un conjunto de datos y desplazamos cada uno por la misma cantidad de unidades, obtenemos una muestra que tiene un coeficiente de variación distinto a la muestra original.

Lo anterior nos indica que, si bien el coeficiente de variación nos sirve para comparar la dispersión entre dos muestras que pueden tener distintas unidades de medida, hay que tener cuidado con la forma en que estas unidades se relacionan.

Curso: Probabilidad y estadística descriptiva
Unidad 2 : Media muestral, dispersión y correlación
Tema: Medidas de tendencia central y de dispersión
Contenido: Propiedades de las medidas de tendencia central y de dispersión

SÍNTESIS

- Si a todos los datos de una muestra se les suma una misma cantidad a (o se le resta una misma cantidad a:

    - la media muestral de los datos desplazados es la media de los datos originales más a(menos a).
    - la desviación estándar de estos nuevos datos es la misma que la de los datos originales.

- Si todos los datos de una muestra se multiplican por una misma cantidad a:

    - la media muestral de estos nuevos datos es la media de los datos originales multiplicada por a.
    - la desviación estándar de estos nuevos datos es la desviación estándar de los datos originales multiplicada por$a$.

- Las propiedades son válidas tanto para datos provenientes de una muestra como para cualquier otro conjunto de datos.

- El coeficiente de variación de un conjunto de datos es una medida de dispersión que se calcula dividiendo la desviación estándar por su media muestral. El coeficiente de variación no tiene unidad de medida.

- Si dos conjuntos de datos tienen la misma desviación estándar, el que tenga una mayor media muestral tendrá un menor coeficiente de variación que la otra muestra.

- Si multiplicamos cada dato de una muestra por una constante, la muestra resultante tendrá un coeficiente de variación igual al de la muestra original.

- Si a cada dato de una muestra le sumamos un valor constante, la muestra resultante tendrá un coeficiente de variación distinto al de la muestra original.