# Estadísticos de dispersión

## Dispersión de una distribución

Las medidas de posición y los promedios que hemos visto en el capítulo anterior, nos dan una síntesis de toda la distribución estadística en un único valor que la representa globalmente, e indica la posición de su "centro". Esto permite hacer de forma sencilla una primera comparación entre distintas poblaciones o muestras. Sin embargo, en muchas ocasiones esta información es insuficiente para llegar a conclusiones válidas. De hecho, dos distribuciones que tengan un mismo valor promedio, pueden diferir muy substancialmente entre si, por ello si atendemos en la comparación solo a los valores promedio podemos cometer graves errores.

El siguiente paso que debe darse en la descripción sintética de la distribución de una variable estadística, es calcular una medida de dispersión. Se llama dispersión de los datos a la variabilidad que existe entre ellos, o dicho de otra forma, al grado en que los valores de la variable estadística tienden a extenderse alrededor del centro o promedio de la distribución.

Las medidas de dispersión serán unas expresiones que para cada distribución nos proporcionarán un número, que si es pequeño, nos indicará que existe poca variabilidad, es decir, que los datos están muy concentrados alrededor del promedio y viceversa, si el número es grande indicará mucha variabilidad, es decir que los valores se hallan muy dispersos con respecto al valor central.. Para calcular estas medidas nos basaremos en las diferencias y desviaciones de los valores que definimos a continuación.

A la diferencia entre un valor de la variable $x_i$ y un promedio P, se le llama diferencia a P y se escribe $( x_i - P )$. Cuando se toma el valor absoluto de esta diferencia $| x_i - P |$ se le denomina desviación, o más precisamente, desviación absoluta a P.

Si consideramos todos los valores de la variable estadística, las operaciones anteriores proporcionan una distribución de las diferencias o desviaciones. La mayor parte de las medidas de dispersión son simplemente un promedio de estas distribuciones de diferencias o desviaciones, o dicho de otro modo, cualquier promedio de la distribución de diferencias o desviaciones constituye una medida de dispersión de la distribución original.

## Algunas medidas de dispersión

### Rangos o recorridos

Se llama rango o recorrido de un conjunto de valores a la diferencia entre el mayor y el menor de todos ellos. Si una variable toma k valores diferentes y los ordenamos de forma creciente : $x^1, x^2, ... x^k$ el rango vendrá dado por $x^k - x^1$.

El rango o recorrido es muy sensible a las fluctuaciones del muestreo, especialmente a la aparición de valores extremos. Esto puede ser paliado utilizando los

rangos entre cuantiles. Así, la longitud del intervalo intercuartílico $Q_3 - Q_1$, contiene el 50% central de la población, ya que dicho intervalo deja a la izquierda el 25% inferior y a la derecha el 25% superior. Puede utilizarse también el rango interdecílico $D_9 - D_1$ que será la longitud del intervalo que contiene el 80% central de la población.

Otra medida del mismo tipo es el rango semiintercuartílico o desviación cuartílica que viene dado por la expresión:

$$\frac{Q_3 - Q_1}{2}$$

Ejemplo: Si las puntuaciones de ocho alumnos una prueba han sido las siguientes: 6, 7, 4, 3, 5, 4, 6, 5 calcule las anteriores medidas de dispersión de esta variable.

En primer lugar, ordenamos los datos en orden creciente, obteniendo la siguiente sucesión: 3, 4, 4, 5, 5, 6, 6, 7. El mayor valor es 7 y el menor 3 en consecuencia el rango será: R = 7 - 3 = 4.

El primer cuartil será el valor que deje por debajo de si una cuarta parte de los individuos, es decir dos individuos, luego $Q_1 = 4$. El tercer cuartil será el valor que deje por debajo de si, las tres cuartas partes de los sujetos, es decir seis individuos, luego $Q_3 = 6$. Por tanto el rango intercuartílico vale 6 - 4 = 2.

De forma inmediata tenemos que el rango semiintercuartílico valdrá 1.

Las anteriores medidas de dispersión utilizan los valores solo por su orden y no por su valor, en consecuencia no aprovechan toda la información existente en la población, o en la muestra. Las restantes medidas que vamos a ver si utilizaran los valores de la variable en su cálculo.

**Varianza y desviación típica**

De los diferentes índices de dispersión, la varianza y la medida asociada a ella, la desviación típica, son las más utilizadas. Esto es debido a sus propiedades algebraicas y al hecho de que la varianza o desviación típica, es uno de los parámetros de una distribución teórica tan importante como la Normal, alrededor de la cual se ha construido una gran parte de la Estadística.

Dado un conjunto de observaciones $x_1, x_2, \ldots x_n$ que tienen como media aritmética $\bar{x}$, se define la varianza, que denotaremos por $S^2$, como la media de los cuadrados de las diferencias a la media, es decir:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

Si los datos vienen agrupados en una tabla, cada diferencia $(x_i - \bar{x})$ ocurrirá $n_i$ veces y en consecuencia la expresión de la varianza será:

$$S^2 = \frac{\sum_{i=1}^{k} n_i \cdot (x_i - \bar{x})^2}{n}$$

En razón de su definición, la varianza se expresa en el cuadrado de las unidades en que se mide la variable, por ello aprovechando que su raíz cuadrada existe siempre por ser un numero positivo, se define la desviación típica, S, como la raíz cuadrada de la varianza, obteniéndose así una medida que se expresa en las mismas unidades que la variable. Por consiguiente, las expresiones de la desviación típica serán:

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \quad \text{y} \quad S = \sqrt{\frac{\sum_{i=1}^{k} n_i \cdot (x_i - \bar{x})^2}{n}}$$

Ejemplo: Volviendo a utilizar las calificaciones de los ocho alumnos, tendremos los siguientes cálculos:

| $x_i$ | 6 | 7 | 4 | 3 | 5 | 4 | 6 | 5 |
|---|---|---|---|---|---|---|---|---|
| $x_i - \bar{x}$ | 1 | 2 | -1 | -2 | 0 | -1 | 1 | 0 |
| $(x_i - \bar{x})^2$ | 1 | 4 | 1 | 4 | 0 | 1 | 1 | 0 |

Por consiguiente, la varianza $S^2$ será igual a 12/8 = 1,5. y la desviación típica S = 1,22.

Si estos mismos datos apareciesen tabulados, las operaciones serían:

| $x_i$ | $n_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $n_i \cdot (x_i - \bar{x})^2$ |
|---|---|---|---|---|
| 3 | 1 | -2 | 4 | 4 |
| 4 | 2 | -1 | 1 | 2 |
| 5 | 2 | 0 | 0 | 0 |
| 6 | 2 | 1 | 1 | 2 |
| 7 | 1 | 2 | 4 | 4 |
| Total | 8 | | | 12 |

Igual que antes, la varianza será 12/8 = 1,5 y la desviación típica 1,22.

Propiedades

Respecto a las propiedades de estas dos medidas citaremos en primer lugar que tanto la varianza como la desviación típica, son invariantes frente a cambios de origen. Es decir, si a todos los valores de la variable les restamos un mismo número, la varianza de los valores resultantes coincide con la de los originales.

En segundo lugar, mencionaremos que la varianza y desviación típica se ven afectadas por los cambios de escala. Si multiplicamos o dividimos todos los valores de una variable por una misma cantidad, la varianza de los valores resultantes es igual a la varianza de los valores originales multiplicada o dividida por el cuadrado de la citada

cantidad. En razón de su definición, resultará que la desviación típica de los valores transformados será igual a la desviación típica de los valores originales multiplicada o dividida por la cantidad mencionada.

Coeficiente de variación

La anterior propiedad nos indica que la varianza y la desviación típica, dependen de la unidad de medida empleada. Por ello no son estadísticos convenientes a la hora de comparar la dispersión de dos o más poblaciones. Para subsanar estas dificultades se define una nueva medida de dispersión, el coeficiente de variación, que es un número sin dimensiones, es decir que no depende de la unidad empleada para medir las variables. Esta medida será la que permita dilucidar, entre dos poblaciones, cual es la que presenta mayor dispersión. Se define como el cociente entre la desviación típica y la media de la distribución, aunque algunos autores multiplican este cociente por 100, es decir:

$$CV = \frac{S}{\bar{x}} \qquad \text{o bien} \qquad CV = \frac{S}{\bar{x}} \cdot 100$$

**Cálculo de la varianza**

Aunque la varianza puede calcularse mediante la formula de su definición, como hicimos en ejemplo correspondiente, en general dicho cálculo resulta laborioso, pues la mayoría de las veces la media no será un valor entero y en consecuencia las diferencias presentarán decimales, al tener que elevar al cuadrado números decimales, se dificulta el cálculo al tiempo que se pierde precisión por los imprescindibles redondeos. Por ello, la mayoría de las veces es preferible utilizar las expresiones siguientes:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 \qquad \text{o bien} \qquad S^2 = \frac{\sum_{i=1}^{k} n_i \cdot x_i^2}{n} - \bar{x}^2$$

Veamos que estas expresiones son equivalentes a las dadas inicialmente

$$S^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n} = \frac{\sum_{i=1}^{n} \left(x_i^2 + \bar{x}^2 - 2x_i\bar{x}\right)}{n} = \frac{\sum x_i^2}{n} + \frac{\sum \bar{x}^2}{n} - \frac{\sum 2x_i\bar{x}}{n} =$$

$$= \frac{\sum x_i^2}{n} + \frac{n\bar{x}^2}{n} - 2\bar{x}\frac{\sum x_i}{n} = \frac{\sum x_i^2}{n} + \bar{x}^2 - 2\bar{x}^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

Ejemplo: El cuadro siguiente da la distribución de las calificaciones de 40 alumnos y las operaciones necesarias para el cálculo de la varianza.

| $x_i$ | $n_i$ | $c_i$ | $n_i \cdot c_i$ | $n_i \cdot c_i^2$ |
|---|---|---|---|---|
| 0 - 2 | 2 | 1 | 2 | 2 |
| 2 - 4 | 7 | 3 | 21 | 63 |
| 4 - 6 | 16 | 5 | 80 | 400 |
| 6 - 8 | 10 | 7 | 70 | 490 |
| 8 - 10 | 5 | 9 | 45 | 405 |
| Totales | 40 | | 218 | 1360 |

$$\bar{x} = \frac{218}{40} = 5,45$$

$$S^2 = \frac{1360}{40} - (5,45)^2 = 34 - 29,7 = 4,3$$

**Puntuaciones diferenciales y típicas**

Como ya mencionamos en el capítulo anterior, al referirnos a las escalas de cuantiles, las puntuaciones que se obtienen directamente, al aplicar un test psicológico, una prueba de rendimiento, etc. carecen de un significado evidente, pues aunque dependen del rasgo o la habilidad del sujeto, también dependen de la longitud y las características que componen la prueba, sin que se pueda, de forma fácil, distinguirse entre ambos factores.

Por otra parte, comparar las puntuaciones de un sujeto en dos pruebas, que tienen distribuciones diferentes, no puede hacerse de forma inmediata. Como tampoco pueden compararse los resultados de dos sujetos, cuando han sido evaluados con instrumentos que aunque midan el mismo concepto, presentan distribuciones diferentes.

Un primer paso en el sentido de hacer interpretables las puntuaciones, es el recurso a las puntuaciones diferenciales. Si a las puntuaciones obtenidas directamente de la aplicación del test o prueba las denominamos puntuaciones directas X, las puntuaciones diferenciales de los sujetos $x$, serán sus puntuaciones directas X menos la media:

$$x = X - \bar{X}$$

Volviendo al ejemplo de las puntuaciones de 146 adolescentes varones en la escala de Ansiedad Estado. Si sabemos que la media es 22, tendremos que el sujeto que había obtenido 30 puntos tendrá una puntuación diferencial de 8 y al que tenía una puntuación de 18, le corresponderá una puntuación diferencial de -4. Así sabemos que el primero, está ocho puntos por encima de la media de todos los sujetos y el segundo cuatro puntos por debajo de la media del grupo. Esto ya nos da una indicación acerca de sus niveles de ansiedad, si son altos o bajos, pero todavía no es suficiente, ya que esas distancias a la media, 8 y -4, no están medidas en unidades interpretables. Para solucionar esta situación se recurre a las puntuaciones típicas.

A partir de las puntuaciones directas, X, se definen las puntuaciones típicas, z, como el resultado de dividir las puntuaciones diferenciales por la desviación típica.

$$z = \frac{X - \overline{X}}{S}$$

Entonces, sabiendo que la desviación típica de la distribución de los 146 adolescentes, en la escala de Ansiedad Estado, era 11 tendremos que sus puntuaciones típicas serán:

$$z = \frac{30 - 22}{11} = 0,73$$

$$z = \frac{18 - 22}{11} = -0,36$$

Estos son unos números sin dimensión que nos indican las distancias a la media de la distribución, medidas en unidades de desviación típica, lo cual , en combinación con ciertos resultados teóricos, hace que sean fácilmente interpretables.

Debido a las propiedades algebraicas de la media y desviación típica, la distribución de las puntuaciones típicas de cualquier prueba, tiene de media cero y desviación típica uno. Esto hace que dos puntuaciones típicas puedan ser comparadas, resolviendo así los problemas que nos planteábamos inicialmente.

Si en la expresión que define las puntuaciones típicas, despejamos las puntuaciones directas obtendremos la regla para calcular la puntuación directa que corresponde a una puntuación típica determinada.

$$X = S \cdot z + \overline{X}$$

Por ejemplo, ¿ Qué puntuación corresponderá en la escala de Ansiedad Estado a una puntuación típica de 2 ?

$$X = 11 \cdot 2 + 22 = 44$$

A pesar de que las puntuaciones típicas resuelven nuestros problemas de interpretación y comparación de puntuaciones, tienen la pequeña dificultad de requerir el manejo de números negativos y fracciones decimales. Por ello, a veces, se prefieren otras puntuaciones estándar, es decir con una media y desviación típica fijadas de antemano, en que los sujetos aparezcan calificados con valores enteros positivos. Por ejemplo, las calificaciones T, cuya media se fija en 50 y su desviación típica en 10.

Sin embargo, las puntuaciones típicas z son la base para el cálculo de cualquier puntuación estándar. Para una puntuación estándar P de media m y desviación típica d, las puntuaciones de los sujetos se obtienen mediante la expresión:

$$P = d \cdot z + m$$

Ejemplo: Calcular la calificación T de un individuo que ha obtenido 7 puntos en un examen, en el cual la media fue 5,5 y la desviación típica 2. En primer lugar calculamos la puntuación típica del alumno:

$$z = \frac{7-5,5}{2} = 0,75$$

y a continuación la transformamos en calificación T.

$$T = 10 \cdot 0,75 + 50 = 57,5$$