

Metodología de la investigación aplicada a la dermatología

Análisis de los datos

María Luisa Peralta*

En el artículo anterior se explicó cómo se describen los datos de forma tabular y gráfica, ahora se pasará a la descripción de los mismos en forma numérica. Para esto, se utilizan medidas de resumen que son de dos tipos: medidas de tendencia central y medidas de dispersión.

Las medidas de tendencia central determinan cuál es el valor promedio de toda la distribución y dónde se ubica, corresponde a lo que comúnmente se llama promedio o media aritmética. La media aritmética se obtiene al sumar las observaciones y el resultado de esta suma se divide entre el número de observaciones; por ejemplo, el promedio de calificaciones de 10 residentes de dermatología fue de 7.6, esta media nos indica dónde se encuentra el centro de la distribución en el continuo de calificaciones posibles (de 0 al 10). Sin embargo, no sólo se requiere establecer dónde se ubica el valor central para describir una distribución de frecuencias, sino también informar el resto de los valores que tan lejos se encuentran de este punto. Para ello se utilizan las medidas de dispersión, como la desviación estándar, cuyo cálculo se efectúa con la siguiente fórmula:

$$DE = \frac{\sum(x-\mu)^2}{n-1}$$

En realidad, no es necesario el cálculo manual, ya que tanto el promedio como la desviación estándar (DE) pueden obtenerse con una calculadora y las pruebas estadísticas más usuales en programas de computo apropiados. Para ello, se necesita restar a cada valor su media (distancia de cada valor a la

media); sin embargo, el objetivo de esta medida no puede calcularse de esta forma, ya que como la dispersión es de cada lado de la media su suma es igual a cero. Las distancias deben elevarse al cuadrado, a fin de obtener números que sean factibles de ser sumados. Al efectuar esta suma y dividirla entre el número de datos (para saber la distancia promedio de los datos a la media), se obtiene lo que se conoce como variancia (otra medida de dispersión). La variancia expresa la dispersión de datos en relación con la media, pero en unidades cuadradas, lo cual hace difícil su interpretación. En este caso, los residentes obtuvieron un promedio de calificación de 7.6 con una variancia de 2.6, es mejor un promedio de 7.6 ± 1.6 (al utilizar la desviación estándar se sabe que 68% de las calificaciones se encuentran entre 6 y 9.2). La variancia sólo es un paso necesario para obtener la desviación estándar (cuadro 1).

Cuadro 1. Principales medidas de resumen:

<i>Medidas de posición</i>	<i>Medidas de dispersión</i>
<ul style="list-style-type: none"> • Media • Mediana • Moda 	<ul style="list-style-type: none"> • Amplitud • Variancia • Desviación estándar • Amplitud o rango intercuartílico

Existen otros promedios de uso frecuente: la mediana y la moda. Con el fin de calcular la mediana, deben ordenarse los valores de forma ascendente; el dato que se encuentra en el centro es la mediana. Si son dos se calcula el promedio de éstos y ésta es la mediana. La quinta y sexta calificación son los dos datos centrales $(8+8)/2 = 8$. Si el 10 no existiera son 9 datos y la mediana es 8, que en este caso coincide. La mediana es más estable que la media, puesto que no se afecta por valores extremos, pero como se calcula al utilizar sólo el lugar de los datos y no el valor de los mismos, proporciona menos información (cuadro 2).

* Médica familiar y maestra en ciencias.

Correspondencia: Dra. María Luisa Peralta. Av. Vértiz 464, colonia Buenos Aires, CP 06780, México, D.F.
E-mail: luisaperalta@netscape.net

Cuadro 2. Cálculo de la desviación estándar de la calificación de 10 residentes de dermatología

Nombre	Calificación (X)	Calif-prom (X-μ)	(Calif-prom) ² (X-μ) ²
MRL	10	2.4	5.76
GFA	9	1.4	1.96
VBA	9	1.4	1.96
ASE	8	0.4	0.16
ADE	8	0.4	0.16
KII	8	0.4	0.16
GHO	7	-0.6	0.36
LHH	7	-0.6	0.36
LPP	6	-1.6	2.56
MGJ	4	-3.6	12.96
Total	76	Σ = 0	26.4

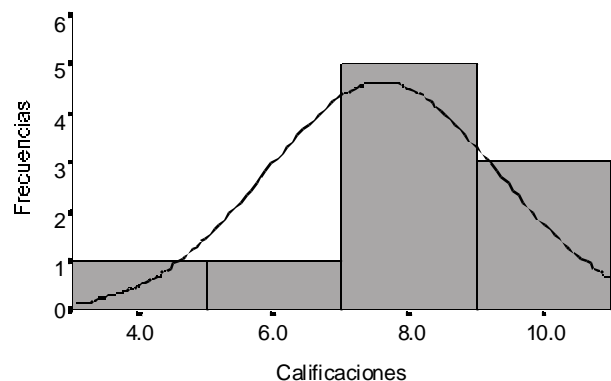
Existen dos parámetros íntimamente relacionados para decidir cuál es la medida de tendencia central que debe utilizarse, la cual determinará la medida de dispersión adecuada y el resto del análisis estadístico, es decir, la naturaleza de la variable en estudio y la forma de la distribución.

Con naturaleza de la variable nos referimos a su escala de medición, si es de razón se utiliza la medida de tendencia central más potente: la media aritmética; ésta es la que proporciona más información y permite hacer todos los cálculos matemáticos. Al calcular la media se utilizan los valores de todas las observaciones, por ende, es la que da más información. No obstante, tiene un defecto, cuando existen valores extremos (muy grandes o viceversa) se desvía y no muestra el dato central, entonces ocasiona una distribución sesgada. En este caso, la más apropiada es la mediana.

Por lo que se refiere a la distribución, es necesario conocer la curva normal o campana de Gauss. Ésta es una representación gráfica de una distribución de frecuencias teórica, que se caracteriza por ser simétrica, mesocúrtica y asintótica. Es simétrica cuando existe la misma cantidad de valores agrupados a uno y otro lado de la media. La curtosis se refiere al grado de elevación de la curva (leptocúrtica: más alta que la normal, platocúrtica: más plana que la normal). Las colas de la curva normal nunca tocan el eje X, por esto se le denomina asintótica. El sesgo y la curtosis de la curva normal son de 0, puesto que es mesocúrtica y simétrica; sin embargo, puede decirse que una distribución de frecuencias se aproxima a la normal

cuando tiene un sesgo de -0.5 a +0.5 y una curtosis de -2 a +2. A medida que la distribución de frecuencias obtenida se aproxime a esta distribución teórica, podrá utilizarse el promedio, la desviación estándar y la estadística paramétrica; de lo contrario, la mejor alternativa para describir los datos son la mediana y el rango intercuartílico (RI), o los centiles y la estadística no paramétrica. Estos últimos son específicos para el manejo de las variables de tipo cualitativo que se miden en escala ordinal o de razón, pero que tienen una distribución diferente a la normal. En el caso de las variables cualitativas que se miden en escala nominal, las frecuencias y los porcentajes son más adecuados.

La forma de la distribución de frecuencias se relaciona con el tamaño de la muestra, de manera que entre menos datos sean, hay mayor probabilidad de que la distribución sea diferente a la normal. Aunque no existe un número específico que asegure la normalidad, con menos de 30 la posibilidad de que se aparte de lo normal es muy alta. La magnitud de la dispersión de los datos también influye en la forma de la distribución, a mayor dispersión (mayor desviación estándar) mayor variabilidad de los datos; por tanto, es más probable que se aleje de lo normal. Son pocos los datos con respecto a las calificaciones, (n = 10) debido a que la dispersión es pequeña (DE = 1.7). A pesar de que casi todas se encuentran más próximas al diez que al cero, hay una calificación extrema que es cuatro, lo que ocasiona que el promedio descienda y se aparte de la mediana (promedio 7.6, mediana 8). La curtosis es de 1.1 y el sesgo de -0.8. A

**Figura 1.** Representación gráfica de la distribución de frecuencias de las calificaciones de 10 residentes de dermatología.

pesar de ser una muestra pequeña, el sesgo no es muy pronunciado y tiende a neutralizarse debido a que la calificación extrema es contraria a la mayor parte de los datos; así, la media y la mediana son muy próximas. El promedio de calificación de los residentes fue de 8 ± 1.1 (md \pm RI), lo que significa que 50% de las calificaciones centrales de los residentes se encontró entre 6.9 y 9.1. El rango intercuartílico se calcula al dividir entre dos la diferencia del percentil 75 y 25. Los percentiles, al igual que el resto de los estadísticos tratados en este artículo, pueden obtenerse con el programa SPSS, es decir, análisis, estadísticas

descriptivas, frecuencias y estadísticas. La curva se describe como discretamente sesgada a la izquierda (la cola es la que determina la dirección del sesgo en la gráfica y de forma numérica por el signo) y ligeramente leptocúrtica (figura 1).

BIBLIOGRAFÍA

1. Argimon PJ, Jiménez VJ. Métodos de investigación clínica y epidemiológica. 2ª ed. Barcelona: Harcourt, 2000.
2. Christensen HB. Estadística paso a paso. 3ª ed. México: Trillas, 1990.
3. Dauson-Saunders B, Trapp R. Bioestadística médica. 2ª ed. México: Manual Moderno, 1997.