
ACERCA DE LA MODA

Eugenio Saavedra Gallardo.

RESUMEN. Después de hacer una breve síntesis sobre la medida de tendencia central moda, en datos no agrupados, se construye la fórmula para la moda en datos agrupados. Esta construcción se realiza a través de una relación de proporcionalidad, que se deriva como consecuencia de la búsqueda de un punto, cuya abscisa es la moda.

ABSTRACT. After a brief synthesis of the definition and main properties of the mode (a measure of central tendency) for ungrouped data, we derive a formula for it in the case of grouped data. Our construction is based on a proportionality relation, that is derived as a consequence of searching for a point whose abscissa is the mode.

§1. Introducción

A lo largo de varios años en que nuestra facultad ha realizado perfeccionamientos para docentes de nuestro sistema escolar (en que han participado alrededor de 700 docentes), hemos observado cómo la gran mayoría de ellos aplican a “ciegas” fórmulas utilizadas en estadística descriptiva, por ejemplo, el caso de los percentiles o el de la moda para datos agrupados en intervalos.

Probablemente esta “ceguera” es producto de la forma en que se presenta, al menos en el sistema escolar, la estadística descriptiva, generalmente solo a través de fórmulas. Caso particular de lo anterior es el de la moda para datos agrupados, la cual puede encontrarse en textos escolares (por ejemplo, las referencias (Bennett, 2015) y (León, 2012)), o en otros libros (por ejemplo las referencias (Bacchini, 2018) y (Salazar, 2018)), y en muchos otros sitios, entre los cuales, solo por nombrar algunos, están las referencias (Farigua, 2016; Sectormatematica, s.f.; Superprof, s.f.; Wiki2, s.f.). En todos ellos, lo que se entrega es solo una fórmula para su cálculo, junto a una identificación de los símbolos que componen dicha fórmula. Sin embargo es preciso señalar que sí existen algunos sitios web donde se muestra,

Palabras clave: Medidas de tendencia central, moda, datos agrupados.

Keywords: Measures of central tendency, mode, grouped data.

en forma bastante resumida, la construcción de dicha fórmula (entre otras, las referencias ([Firmfunda, s.f.](#); [Topperlearning, s.f.](#); [Math.Stackexchange, s.f.](#))).

Motivados por lo anterior se escribe este artículo, con el objetivo de construir detalladamente la fórmula para calcular la moda en datos agrupados, y de esta manera mostrar de forma pormenorizada qué es lo que se pretende capturar con ella.

La construcción se realiza por medio de una relación de proporcionalidad que involucra tanto la frecuencia del intervalo modal como la de los intervalos contiguos. Esta relación de proporcionalidad se deriva como consecuencia de la búsqueda de un punto del cual la moda es su abscisa, además la búsqueda de este punto se hace tanto desde una mirada algebraica (intersección de dos rectas en el plano), cómo desde una mirada geométrica (semejanza de triángulos).

El artículo comienza con la presentación de las 3 medidas comunmente utilizadas para describir la idea de centro de un conjunto de datos (media o promedio, moda y mediana), para luego analizar brevemente la moda, tanto para datos cuantitativos (no agrupados) como para datos cualitativos (categorías). Seguidamente se entrega la fórmula para el cálculo de la moda en datos agrupados, mostrando su operatoria a través de un ejemplo. Posteriormente, se hace la construcción y el análisis de la fórmula para la moda en datos agrupados, para luego reinterpretar la forma en que esta se expresa.

§2. La idea de centro

Comúnmente 3 medidas son usadas para describir el “centro” o “localización central” de un conjunto de datos (a este conjunto también se le llama datos muestrales o muestra):

- la moda,
- la mediana,
- la media (o promedio).

Cada una de estas medidas interpreta la idea de “centro” en diferentes maneras. La moda interpreta el significado de “centro” como el valor que ocurre más a menudo en el conjunto de datos analizados, en cambio la mediana interpreta el significado de “centro” como el valor que divide al conjunto de datos, después de ordenarlos en forma ascendente, en dos partes, de modo que al menos el 50 % de los datos son menores o iguales a la mediana y al menos el 50 % de los datos son mayores o iguales a esta. Por último el promedio interpreta la idea de “centro” como una especie de repartición “equitativa” en el siguiente sentido. Tres amigos coleccionan poster de su cantante preferido, ellos tienen 19, 31 y 13 poster respectivamente. Los amigos deciden repartirse los poster de modo que cada uno quede con la misma cantidad, ¿con cuántos poster queda cada uno? La respuesta es 21, que corresponde al promedio de los valores 19, 31 y 13.

La elección de una apropiada medida para cualquier conjunto de datos, dependerá de al menos dos factores:

- Los aspectos de la “localización central” que pretenden capturarse.
- El tipo de datos de que se disponga, es decir, si los datos son cuantitativos (numéricos) o cualitativos (categorías).

La segunda consideración es importantísima, porque, dependiendo del tipo de datos, será posible calcular todas o sólo algunas de las medidas de tendencia central recién mencionadas:

- si los datos son cuantitativos entonces la moda, mediana y media pueden ser calculados,
- si los datos son cualitativos, la única medida de localización central que puede ser calculada es la moda, ya que por el significado de “centro” de la mediana y del promedio, para su cálculo es necesario que operaciones algebraicas con los datos sean posibles.

§3. La moda para datos cuantitativos no agrupados

Como dijimos anteriormente, la moda interpreta la idea de “centro” como el valor que ocurre más a menudo en el conjunto de datos analizados. El valor de la moda se puede calcular de la siguiente forma:

- arreglar los datos en forma creciente,
- contar el número de veces que cada valor ocurre,
- identificar el valor que ocurre con mayor frecuencia. Este valor es la moda.

Cuando existan solo dos datos con mayor frecuencia, se dice que la distribución de los datos es bimodal, o sea, tiene dos modas. En el caso en que al menos tres datos tengan mayor frecuencia, se dice que la distribución de los datos es multimodal (tiene al menos 3 modas). En particular, si todos los datos tienen igual frecuencia, entonces la distribución es multimodal.

Propiedades de la moda

- El cálculo de la moda no usa toda la información en la muestra, es decir, no involucra todos los valores obtenidos en la muestra, sólo usa el valor que ocurre más a menudo. Este hecho podría considerarse como una seria debilidad si pensamos que toda la información disponible en la muestra debería ser usada.

Debido a lo anterior, la moda podría no ser muy descriptiva de lo que ocurre en la muestra. Por ejemplo, si las notas (en escala de 0 a 100) de un alumno en la asignatura de matemática fuesen:

88, 83, 40, 84, 35, 86, 95, 40, 92, 85,

que ordenadas de menor a mayor resultan

35, 40, 40, 83, 84, 85, 86, 88, 92, 95,

obtendríamos que la moda es la nota 40, por lo que juzgar el rendimiento de este alumno en matemática, por medio de la moda, parece poco adecuado.

- La moda es una medida que es sensible a pequeños cambios de los valores muestrales.
- La moda no es particularmente afectada por valores extremos en la muestra, es decir, por valores que resultan mucho más grandes o mucho más pequeños que el resto de los datos.
- La moda es siempre igual a uno de los valores presentes en la muestra (en el caso de datos no agrupados).
- No hay un orden (estricto) predeterminado entre la moda y las otras dos medidas de tendencia central. En efecto, si denotamos por M_d a la mediana, M_0 a la moda y \bar{x} al promedio, entonces puede ocurrir que:

$M_d < M_0 < \bar{x}$, si los datos son 0, 1, 2, 3, 4, 5, 6, 6, 98, 99, 100,

$M_d < \bar{x} < M_0$, si los datos son -1, 0, 1, 5, 5,

$M_0 < \bar{x} < M_d$, si los datos son -3, -3, -2, 0, 1, 2, 3,

$M_0 < M_d < \bar{x}$, si los datos son -1, -1, 0, 3, 4,

$\bar{x} < M_0 < M_d$, si los datos son -100, -4, -4, 1, 2, 3, 4,

$\bar{x} < M_d < M_0$, si los datos son -3, -2, 0, 1, 1,

$\bar{x} = M_0 = M_d$, si los datos son -2, 0, 0, 2.

§4. La moda para datos cualitativos (categorías)

En este caso, es decir, cuando los datos son cualitativos, la moda es la única medida de tendencia central que se puede calcular para este tipo de datos. El procedimiento es el mismo que en el caso de datos cuantitativos (numéricos). La moda es(son) la(s) categoría(s) que ocurre(n) más frecuentemente en el conjunto de datos analizados (muestra).

Por ejemplo, una Compañía de telefonía móvil realiza una encuesta a 10 de sus clientes, acerca de su servicio, entregándole como opciones: muy satisfecho (MS); satisfecho (S); insatisfecho (I) y muy insatisfecho (MI). La Tabla 1 más abajo resume la información recogida.

En este caso la moda resulta ser la categoría “muy satisfecho”, es decir, lo que más frecuentemente ocurre (en los datos de la muestra) es que el cliente se encuentre muy satisfecho con el servicio de telefonía móvil que le ofrece la compañía.

Categoría	Frecuencia
Muy insatisfecho	2
Insatisfecho	3
Satisfecho	1
Muy satisfecho	4

Tabla 1

A través de este mismo ejemplo podemos ver que el cálculo de la media no tiene sentido. Si tratáramos de calcular la media, deberíamos sumar los 10 datos, lo cual obviamente no es posible.

Incluso si asignamos valores numéricos a las categorías (esta operación es conocida como codificación), por ejemplo, muy insatisfecho = 0, insatisfecho = 1, satisfecho = 2, muy satisfecho = 3, la media todavía continúa sin sentido. Esto porque los valores numéricos que asignamos a las diferentes categorías, asumen que la diferencia entre la respuesta “satisfecho” y “muy satisfecho” es la misma que la diferencia entre la respuesta “muy insatisfecho” e “insatisfecho”, lo cual en general no corresponde a la realidad. Puede ser que “estar muy satisfecho” sea mucho mejor que “estar satisfecho”.

El mismo argumento anterior se aplica a datos cualitativos ordinales numéricos. Por ejemplo, el test de Apgar, arroja como resultado un número que puede ser 0, 1, . . . , 10. Pero, un recién nacido con puntaje 10 no significa que sea el doble de saludable que uno con puntaje 5.

También, para este tipo de datos no tiene sentido calcular la mediana. Por ejemplo, con los datos de la tabla anterior, la mediana sería el promedio entre I y S, lo cual no tiene sentido.

§5. La moda para datos cuantitativos agrupados

Consideremos un conjunto de datos agrupados, los cuales pueden presentarse a través de una tabla de datos agrupados en intervalos, un histograma o una ojiva. Además, estas presentaciones pueden mostrarse con frecuencias absolutas y/o frecuencias relativas y/o frecuencias relativas porcentuales. Por otra parte, teniendo un tipo de presentación de los datos agrupados, inmediatamente pueden obtenerse las otras dos.

Si llamamos M_0 a la moda para datos agrupados en intervalos (de igual amplitud), entonces M_0 se define como

$$M_0 = \lim_{\text{inf}} + \frac{d_1}{d_1 + d_2} \cdot a \quad \text{ó} \quad M_0 = \lim_{\text{inf}} + \frac{f_M - f_1}{2f_M - f_1 - f_2} \cdot a,$$

donde

- I_M = intervalo con mayor frecuencia (se conoce como intervalo modal),
 \lim_{inf} = límite inferior de I_M ,
 \lim_{sup} = límite superior de I_M ,
 a = amplitud de I_M , es decir, $\lim_{\text{sup}} - \lim_{\text{inf}}$,
 f_M = frecuencia del intervalo I_M ,
 f_1 = frecuencia del intervalo contiguo por la izquierda a I_M ,
 f_2 = frecuencia del intervalo contiguo por la derecha a I_M ,
 $d_1 = f_M - f_1$,
 $d_2 = f_M - f_2$.

Observaciones:

- En la fórmula, cuando se habla de frecuencia, esta puede ser la frecuencia absoluta, la frecuencia relativa o la frecuencia relativa porcentual. Con cualquiera de ellas el valor de M_0 es el mismo.
- Si el intervalo modal se encuentra en el primer intervalo, entonces en la fórmula se considera que $f_1 = 0$ (el intervalo contiguo a la izquierda del intervalo modal tiene frecuencia 0, por ello no aparece). Similarmente, si el intervalo modal se encuentra en el último intervalo, entonces en la fórmula se considera que $f_2 = 0$ (el intervalo contiguo a la derecha del intervalo modal tiene frecuencia 0, por ello no aparece)
- Si hay más de un intervalo modal, entonces se aplica la fórmula a cada uno de estos intervalos, con lo cual se obtienen tantas modas como intervalos modales se tenga.

Por ejemplo, los siguientes datos corresponden al puntaje obtenido por 80 estudiantes de diversos establecimientos escolares en una prueba estandarizada.

Intervalo	Frecuencia Absoluta	Frecuencia Relativa Porcentual
[395 – 455[6	7,5 %
[455 – 515[8	10 %
[515 – 575[10	12,5 %
[575 – 635[22	27,5 %
[635 – 695[16	20 %
[695 – 755[12	15 %
[755 – 815]	6	7,5 %
Total	80	100 %

Tabla 2

A partir de la tabla anterior obtenemos:

Símbolo	Con Frecuencia Absoluta	Con Frecuencia Relativa Porcentual
\lim_{inf}	575	575
\lim_{sup}	635	635
a	60	60
f_M	22	27,5 %
f_1	10	12,5 %
f_2	16	20 %
d_1	12	15 %
d_2	6	7,5 %
$\frac{d_1}{d_1 + d_2}$	$\frac{12}{12 + 6} = \frac{2}{3}$	$\frac{15\%}{15\% + 7,5\%} = \frac{2 \cdot 7,5}{3 \cdot 7,5} = \frac{2}{3}$

Tabla 3

Así,

$$M_0 = \lim_{\text{inf}} + \frac{d_1}{d_1 + d_2} a = 575 + \frac{2}{3} 60 = 615.$$

Si conociéramos explícitamente los 80 datos (por ejemplo, los que se presentan a continuación y que satisfacen las condiciones de la Tabla 2)

395	410	410	410	410	449	460	495	500	505
511	511	511	513	515	515	521	523	530	550
562	570	571	573	575	575	580	581	582	585
585	585	588	590	592	595	598	599	600	615
615	616	620	625	630	631	635	635	639	640
645	648	650	652	655	655	660	660	665	667
680	690	700	708	710	725	725	730	735	738
740	747	750	754	778	780	790	795	800	815

Tabla 4

veríamos que la verdadera moda es 410. El error que cometimos al obtener como moda el valor 615, es el precio que se debe pagar por el desconocimiento de los datos cuando estos se presentan agrupados y no se conocen explícitamente.

§6. Construcción de la Fórmula

Primero, recordemos que la fórmula para M_0 , la moda en datos agrupados, es

$$M_0 = \lim_{\text{inf}} + \frac{d_1}{d_1 + d_2} a.$$

En adelante, para simplificar, usaremos $L_i = \lim_{\text{inf}}$ y $L_s = \lim_{\text{sup}}$.

¿Qué trata de expresar la fórmula de la moda para datos agrupados? Trata de escoger un valor (llamado M_0) dentro del intervalo modal.

¿Con que criterio se escoge M_0 ? El criterio que se utiliza para escoger M_0 , es la comparación entre la frecuencia del intervalo contiguo por la izquierda al intervalo modal, esto es f_1 , y la frecuencia del intervalo contiguo por la derecha, o sea f_2 .

Si f_1 es mayor que f_2 , entonces el punto M_0 se debe “acercar” más al extremo izquierdo del intervalo modal, es decir, a L_i , en cambio, si f_1 es menor que f_2 , el punto M_0 se debe “acercar” más al extremo derecho del intervalo modal, es decir, a L_s .

Para el caso en que las frecuencias de los intervalos contiguos sean iguales, entonces el punto escogido debe corresponder al punto medio del intervalo modal.

¿Cómo decide la fórmula “acercarse” más al extremo izquierdo (respectivamente derecho) del intervalo modal?

La forma como lo hace es escogiendo a M_0 como la abscisa del punto de intersección entre los segmentos de línea punteada que se muestran en el gráfico siguiente.

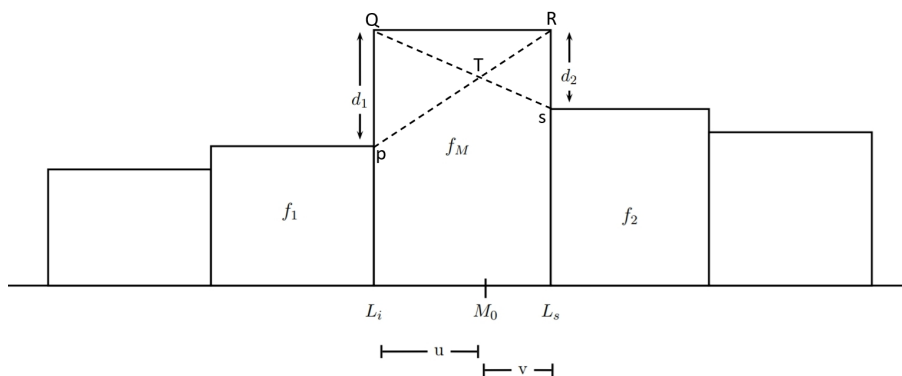


Gráfico 1

Cabe señalar que el gráfico anterior es la representación de un conjunto de datos agrupados en intervalos, en el sentido que en el eje horizontal se ponen los intervalos y la altura de un rectángulo representa la frecuencia correspondiente al intervalo que forma su base.

¿Qué relaciones algebraicas satisface el punto de intersección entre los segmentos de línea punteada?

Como M_0 se escoge de modo que sea la abscisa del punto intersección del segmento \overline{PR} con el segmento \overline{SQ} (ver Gráfico 1), entonces M_0 puede ser determinado encontrando la intersección de la recta que pasa por los puntos $P = (L_i, f_1)$ y $R = (L_s, f_M)$, con la recta que pasa por los puntos $Q = (L_i, f_M)$ y $S = (L_s, f_2)$.

La ecuación de la recta que pasa por los puntos P y R se puede escribir como

$$y = \frac{d_1}{a}x + f_1 - \frac{d_1}{a}L_i.$$

La ecuación de la recta que pasa por los puntos Q y S se puede escribir como

$$y = -\frac{d_2}{a}x + f_2 + \frac{d_2}{a}L_s.$$

Como el punto de intersección de las rectas, que es de la forma (M_0, y_0) , debe satisfacer las ecuaciones que las determinan, entonces M_0 debe satisfacer la relación

$$\frac{d_1}{a}M_0 + f_1 - \frac{d_1}{a}L_i = -\frac{d_2}{a}M_0 + f_2 + \frac{d_2}{a}L_s,$$

esto es,

$$d_1M_0 + af_1 - d_1L_i = -d_2M_0 + af_2 + d_2L_s.$$

Por lo cual, utilizando las relaciones $L_i = -(L_s - L_i) + L_s = -a + L_s$ y $L_s = (L_s - L_i) + L_i = a + L_i$ se tiene que

$$d_1M_0 + af_1 - d_1(-a + L_s) = -d_2M_0 + af_2 + d_2(a + L_i),$$

o sea

$$d_1M_0 + af_1 + d_1a - d_1L_s = -d_2M_0 + af_2 + d_2a + d_2L_i,$$

de donde

$$d_1(M_0 - L_s) = -d_2(M_0 - L_i) - af_1 - ad_1 + af_2 + ad_2.$$

Pero, $d_1 = f_M - f_1$ y $d_2 = f_M - f_2$, por lo cual

$$-af_1 - ad_1 + af_2 + ad_2 = a(-f_1 - (f_M - f_1) + f_2 + (f_M - f_2)) = 0.$$

En consecuencia

$$d_1(M_0 - L_s) = -d_2(M_0 - L_i),$$

esto es,

$$\frac{M_0 - L_i}{L_s - M_0} = \frac{d_1}{d_2}. \quad (*)$$

Finalmente, esta relación de proporcionalidad es la que nos permite determinar M_0 . En efecto, como

$$M_0 - L_i = \frac{d_1}{d_2}(L_s - M_0),$$

entonces

$$M_0 + \frac{d_1}{d_2} M_0 = L_i + \frac{d_1}{d_2} L_s,$$

por lo cual

$$M_0 \frac{d_1 + d_2}{d_2} = L_i + \frac{d_1}{d_2} L_s,$$

esto es

$$M_0 = \frac{d_2}{d_1 + d_2} L_i + \frac{d_2}{d_1 + d_2} \frac{d_1}{d_2} L_s.$$

Por otra parte, $L_s = (L_s - L_i) + L_i = a + L_i$, así

$$\begin{aligned} M_0 &= \frac{d_2}{d_1 + d_2} L_i + \frac{d_1}{d_1 + d_2} (a + L_i) \\ &= \left(\frac{d_2}{d_1 + d_2} + \frac{d_1}{d_1 + d_2} \right) L_i + \frac{d_1}{d_1 + d_2} a \\ &= L_i + \frac{d_1}{d_1 + d_2} a = \lim_{\inf} + \frac{d_1}{d_1 + d_2} a. \end{aligned}$$

Para el caso en que $f_1 = f_2$, esto es, cuando $d_1 = d_2$, la proporción entre d_1 y d_2 es uno, por lo que la proporción entre $(M_0 - L_i)$ y $(L_s - M_0)$ debe ser uno. Para que esto se cumpla, M_0 debe ser el punto medio del intervalo modal.

En efecto, al ocurrir que $d_1 = d_2$, se tiene que $\frac{d_1}{d_1 + d_2} = \frac{1}{2}$, por lo que

$$M_0 = L_i + \frac{d_1}{d_1 + d_2} a = L_i + \frac{1}{2} a = L_i + \frac{1}{2} (L_s - L_i) = \frac{L_i + L_s}{2},$$

que corresponde al punto medio del intervalo modal.

Observación: ¿Qué relaciones geométricas satisface el punto de intersección de los segmentos \overline{PR} y \overline{SQ} ?

El gráfico que se presenta a continuación, corresponde a la “parte superior” del rectángulo que tiene por base al intervalo modal que se muestra en el Gráfico 1, por lo cual PQ corresponde a d_1 , RS corresponde a d_2 y T a la intersección del segmento \overline{PR} con el segmento \overline{SQ} .

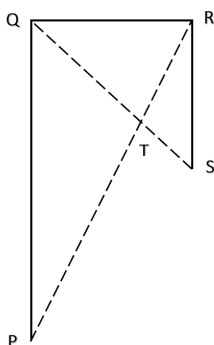


Gráfico 2

Del gráfico anterior se puede observar que, si $f_1 < f_2$, entonces PQ será mayor que RS , por lo que T estará más cerca del segmento \overline{RS} que del segmento \overline{PQ} . Para visualizar lo anterior, basta recordar que las diagonales de un rectángulo se cortan en el centro de este.

En el caso en que $f_1 > f_2$, tendremos un gráfico análogo al Gráfico 2, solo que en este caso PQ será menor que RS , por lo que T estará más cerca del segmento \overline{PQ} que del segmento \overline{RS} .

Ahora, si $f_1 = f_2$, entonces $PQ = RS$, por lo que el cuadrilátero $PQRS$ será un rectángulo. Por esta razón T estará en el centro de este, por tanto T será equidistante de los segmentos \overline{PQ} y \overline{RS} .

También, desde el Gráfico 2, utilizando que $\overline{PQ} \parallel \overline{RS}$ y el criterio de semejanza de triángulos AAA , se tiene que $\triangle PTQ \sim \triangle RTS$, de donde $\frac{PQ}{RS} = \frac{PT}{RT}$.

El gráfico que se presenta ahora, es similar al Gráfico 2, solo que ahora se agregaron unas líneas punteadas auxiliares.

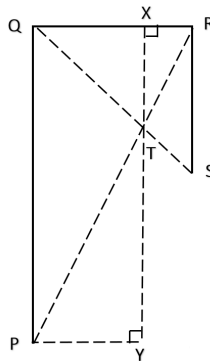


Gráfico 3

Usando nuevamente el criterio de semejanza de triángulos AAA , se deduce que $\triangle PYT \sim \triangle RXT$, por lo que se tiene $\frac{PT}{RT} = \frac{PY}{RX}$.

En consecuencia, $\frac{PQ}{RS} = \frac{PT}{RT} = \frac{PY}{RX}$, de donde $\frac{PQ}{RS} = \frac{PY}{RX}$, o dicho de otra forma,

$$\frac{PY}{PQ} = \frac{RX}{RS}.$$

A partir de los Gráficos 1, 2 y 3, vemos que $PY = u$, $PQ = d_1$, $RX = v$ y $RS = d_2$, por lo que la última expresión puede escribirse de la siguiente manera:

La proporción entre la magnitud del segmento $\overline{L_i M_0}$ y d_1 , es igual a la proporción entre la magnitud del segmento $\overline{M_0 L_s}$ y d_2 .

En otras palabras

$$(M_0 - L_i) : d_1 = (L_s - M_0) : d_2,$$

es decir

$$\frac{M_0 - L_i}{L_s - M_0} = \frac{d_1}{d_2}. \quad (*)$$

O sea, M_0 debe satisfacer la misma relación de proporcionalidad (*), a partir de la cual, anteriormente, obtuvimos la fórmula para M_0 .

Observación: En rigor, la principal razón para escoger a M_0 como la abscisa del punto de intersección entre los segmentos \overline{PR} y \overline{SQ} (ver Gráfico 1), es tratar de capturar el valor donde una curva “suave” (densidad de probabilidad que pudiese ajustarse a los datos agrupados, que están representados en el Gráfico 1) alcanzaría su valor máximo, esto es, capturar la moda de la distribución de probabilidad que se estaría ajustando a los datos.

A modo de ejemplo: El siguiente gráfico muestra los datos asociados a la Tabla 2, nuevamente, en el sentido que la altura de un rectángulo representa la frecuencia correspondiente al intervalo que forma su base.

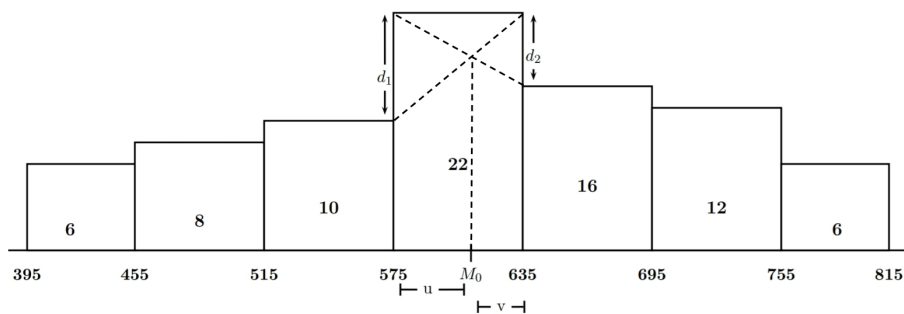


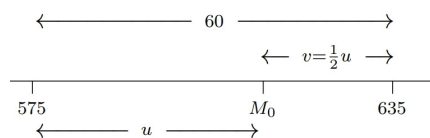
Gráfico 4

En este caso, $u = M_0 - 575$, $v = 635 - M_0$, $d_1 = 12$ y $d_2 = 6$. Como $f_1 = 10$ menor que $f_2 = 16$, entonces M_0 debe estar más “cerca” del extremo derecho del intervalo modal, es decir, más cerca de $L_s = 635$.

Por otra parte, $\frac{d_1}{d_2} = 2$, por lo que $\frac{u}{v}$ debe ser igual a 2 (puesto que la forma como decide la fórmula acercarse a uno u otro extremo del intervalo modal, se traduce en que ambas proporciones sean iguales). Dicho de otra forma, como d_1 es el doble de d_2 , entonces u debe ser el doble que v .

También, $u + v = 60$, o sea, $2v + v = 60$, por lo que $v = \frac{60}{3} = 20$, de donde $u = 40$. Pero, $M_0 = 575 + u$, por lo cual M_0 debe ser 615 (resultado obtenido en la Sección 5).

La figura siguiente ilustra lo anterior.



§7. Otra Aproximación a la Fórmula

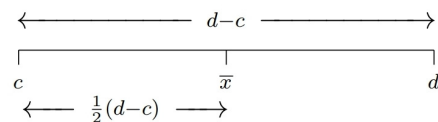
Sean c y d números reales, con $c < d$, entonces el promedio de estos dos valores es $\bar{x} = \frac{c+d}{2}$, el cual puede escribirse como $\bar{x} = \frac{1}{2}c + \frac{1}{2}d$, y corresponde al punto medio del intervalo $]c, d[$.

De forma más general se define el promedio ponderado para c y d como el número real $\bar{x}_p = \alpha c + \beta d$, donde α y β son números reales pertenecientes al intervalo $]0, 1[$, y tal que $\alpha + \beta = 1$.

Un caso particular de promedio ponderado es cuando $\alpha = \beta = \frac{1}{2}$, esto es, el promedio de los valores c y d es un caso particular de promedio ponderado.

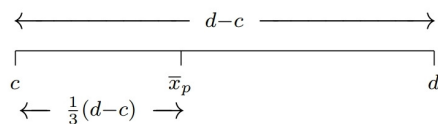
A continuación veamos que \bar{x}_p pertenece al intervalo $]c, d[$, y que \bar{x}_p estará más cerca de c cuando β sea menor que α , más cerca de d cuando β sea mayor que α y al centro del intervalo cuando β sea igual a α . Además veremos que la moda para datos agrupados es un promedio ponderado de los extremos del intervalo modal.

Primeramente la distancia entre c y d es $(d - c)$ y por otra parte $\bar{x} = \frac{1}{2}c + \frac{1}{2}d = c + \frac{1}{2}(d - c)$.



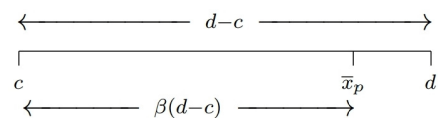
Esto es, la distancia entre c y \bar{x} es la misma que entre \bar{x} y d , en otras palabras \bar{x} es el punto medio del intervalo de extremos c y d .

Ahora sea $\bar{x}_p = \frac{2}{3}c + \frac{1}{3}d = c + \frac{1}{3}(d - c)$.



Entonces la distancia entre c y \bar{x}_p es la mitad de la distancia entre \bar{x}_p y d .

Sea ahora $\bar{x}_p = \alpha c + \beta d$. Como $\alpha + \beta = 1$, entonces $\bar{x}_p = (1 - \beta)c + \beta d$, esto es, $\bar{x}_p = c + \beta(d - c)$.



Por lo tanto la distancia entre c y \bar{x}_p es $\beta(d - c)$ y la distancia entre \bar{x}_p y d es $(1 - \beta)(d - c)$, con $\beta \in]0, 1[$ (en los casos anteriores β era $\frac{1}{2}$ y $\frac{1}{3}$, respectivamente).

Así, en el caso en que $\beta < \frac{1}{2}$, \bar{x}_p estará más cerca de c que de d y si $\beta > \frac{1}{2}$, \bar{x}_p estará más cerca de d que de c .

Ahora reescribamos la fórmula de la moda.

$$M_0 = L_i + \frac{d_1}{d_1+d_2}a = L_i + \frac{d_1}{d_1+d_2}(L_s - L_i) = \frac{d_2}{d_1+d_2}L_i + \frac{d_1}{d_1+d_2}L_s.$$

Entonces M_0 resulta ser el promedio ponderado entre L_i y L_s , con $\alpha = \frac{d_2}{d_1+d_2}$ y $\beta = \frac{d_1}{d_1+d_2}$.

Notar que, en el caso en que el intervalo contiguo por la izquierda al intervalo modal tenga mayor frecuencia que el intervalo contiguo por la derecha, esto es, cuando $f_1 > f_2$, se tendrá que $f_M - f_1 < f_M - f_2$, o sea $d_1 < d_2$. Esta desigualdad es equivalente a $\beta = \frac{d_1}{d_1+d_2} < \frac{1}{2}$. O sea, en este caso la moda M_0 estará más cerca del límite inferior del intervalo modal que del límite superior.

Análogamente, si $f_1 < f_2$, entonces $\beta = \frac{d_1}{d_1+d_2} > \frac{1}{2}$, por lo que M_0 estará más cerca del límite superior del intervalo modal que del límite inferior.

En el ejemplo de la Sección 5, $f_1 = 10$, $f_2 = 16$ y $f_M = 22$, por lo que $d_1 = 12$, $d_2 = 6$ y $\beta = \frac{2}{3}$. En consecuencia,

$$M_0 = \frac{1}{3} \cdot 575 + \frac{2}{3} \cdot 635 = 615.$$

En este caso $f_1 < f_2$, o sea, $\beta > \frac{1}{2}$, por lo cual la moda $M_0 = 615$ está más cerca de 635 que de 575 (la distancia entre la moda y 575 es 40, mientras la distancia entre la moda y 635 es 20).

§8. Comentarios finales

Con esto finalizamos el análisis de la fórmula para calcular la moda desde datos agrupados. Esta fórmula es la que se presenta (sin ninguna explicación, salvo la identificación de cada símbolo que la compone) tanto en algunos textos como en otros recursos pedagógicos que se utilizan en el sistema escolar. Las preguntas que surgen en forma natural respecto a esta fórmula son: ¿valdrá la pena, al menos en el sistema escolar, presentar esta fórmula sin ninguna explicación sobre su construcción? ¿será necesario, al menos para el sistema escolar, que la moda esté más cerca del límite inferior o del límite superior del intervalo modal si este valor puede diferir tanto respecto al verdadero valor de la moda (como en el ejemplo de la Sección 5)?

Agradecimientos. El autor quiere agradecer a los revisores anónimos del artículo, por sus valiosas observaciones y sugerencias, las cuales permitieron mejorar la redacción del mismo.

Bibliografía

- Bacchini, e. a., R. (2018). *Introducción a la probabilidad y a la estadística*. Universidad de Buenos Aires. Descargado de <http://bibliotecadigital.econ.uba.ar/download/libros/>
- Bennett, e. a., J. (2015). *Matemática 8° Básico, Texto del Estudiante*. Edición para el Ministerio de Educación de Chile, Editorial Galileo, Santiago, Chile.
- Farigua, K. (2016). *Propuesta de enseñanza para medidas de tendencia central a través de objetos virtuales de aprendizaje*. Universidad Pedagógica Nacional, Bogota. Descargado de <http://funes.uniandes.edu.co/12013/1/Farigua2016Propuesta.pdf>
- Firmfunda. (s.f.). <https://firmfunda.com/statistics-basics>. (Descargado la fecha de publicación del artículo)
- León, P. e. a. (2012). *Matemática 8° Básico, Texto del Estudiante*. Edición para el Ministerio de Educación de Chile, Editorial MN, Santiago, Chile.
- Math.Stackexchange. (s.f.). <https://math.stackexchange.com/questions/3687290/derivation-of-the-equation-of-mode-for-grouped-data>. (Descargado la fecha de publicación del artículo)
- Saavedra, E. (2020). *Contenidos básicos de estadística y probabilidad, 3^{ra} edición*. Editorial Usach, Santiago, Chile.
- Salazar, S., C. y Del Castillo. (2018). *Fundamentos básicos de estadística*. Universidad Central del Ecuador. Descargado de <https://www.dspace.uce.edu.ec>
- Sectormatematica. (s.f.). <https://sectormatematica.cl/educmedia.htm>. (Descargado la fecha de publicación del artículo)
- Superprof. (s.f.). <https://superprof.es/apuntes/escolar\protect\penalty-\@M/matematicas/estadistica/descriptiva/>. (Descargado la fecha de publicación del artículo)
- Topperlearning. (s.f.). <https://www.topperlearning.com/answer/what-is-the-deivation-of-the-formulae-for-finding-median-and-mode-of-grouped-data-as-listed-in-class-10-mathematics-book/url59dcc>. (Descargado la fecha de publicación del artículo)
- Wiki2. (s.f.). [https://wiki2.org/es/Moda_\(estadistica\)](https://wiki2.org/es/Moda_(estadistica)). (Descargado la fecha de publicación del artículo)

EUGENIO SAAVEDRA GALLARDO.

Departamento de Matemática y C.C. , Universidad de Santiago de Chile

(✉) eugenio.saavedra@usach.cl

Recibido: *17 de abril de 2020.*

Aceptado: *21 de diciembre de 2020.*

Publicado en línea: *5 de abril de 2021.*
