

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

FICHA Nº 17

A. Para el análisis descriptivo de las variables métricas suelen emplearse distintas técnicas. Tal como lo recuerda y desarrolla Cortés (2000:129) en el camino del resumen de la información pueden utilizarse estrategias tales como las tablas de frecuencias simples, proporciones, porcentajes y porcentajes acumulados.

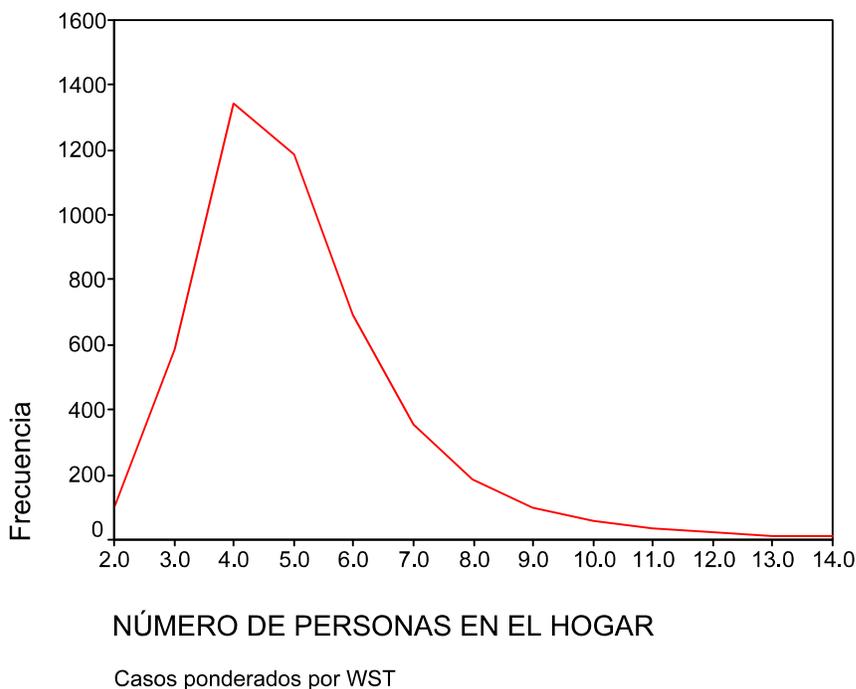
- ❖ Es razonable informar con proporciones la distribución del número de miembros que componen un hogar. Por ejemplo, en el siguiente cuadro, la variable se resume mediante proporciones tomadas de la encuesta sociofamiliar que se aplicó en el marco de la evaluación de aprendizajes en 6to. De Primaria en Uruguay (1999). El resumen que se presenta resulta elocuente en cómo se distribuye la variable entre los hogares. Sin embargo, también se puede observar que un 39% de los hogares tiene 6 y más miembros. No es posible conocer para estos 1960 hogares cuál es el tamaño que tienen en realidad: la última categoría o intervalo resulta muy grande.

	Frecuencias simples	Proporciones
2 personas	101	0,02
3 personas	586	0,12
4 personas	1343	0,27
5 personas	1189	0,24
6 y más personas	1960	0,39
Sin datos	294	0,06
Total	4988	1,00

- ❖ Si por ejemplo, la variable a resumir fuera el ingreso, es claro que la misma conformación de las categorías para luego proceder a porcentualizar la distribución, generaría un problema de análisis. En consecuencia, el resumen de las variables métricas parecería ser más apropiado mediante estadísticos más eficientes: idealmente, quisiéramos tener una única medida que resumiera toda la distribución.

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

- Junto con los estadísticos que se desarrollarán a continuación, es recomendable comenzar a acostumbrarse a representar gráficamente la distribución observada de la variable. Las gráficas nos auxiliarán a visualizar qué tipo de estadísticos va a ser más apropiado para resumir la distribución y qué inconveniente tiene cada tipo. En el ejemplo anterior, la gráfica es la siguiente:



B. Las medidas de tendencia central tienen por objetivo resumir la distribución univariada de datos informando cuál es el valor más representativo de la variable analizada (Ver ficha 10).



Esta familia de estadísticos sólo puede aplicarse a las variables denominadas “métricas” sean de tipo “interval” o “de razón”.

- Establecer la pertinencia del estadístico según el tipo de variables es una responsabilidad del investigador. Los paquetes calculan los estadísticos aunque estos no tengan sentido, como en el caso de las

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

variables nominales pluricotómicas.

- ii) Un caso particular lo constituyen los reactivos que conforman una escala de tipo "Likert" de 5, 6 o 7 puntos. Se supone que el encuestado ubica su respuesta en un continuo (por ejemplo, muy de acuerdo a muy en desacuerdo) pero que se captura en forma discreta a través de una variable ordinal. Se acepta por convención que esta variable puede ser objeto de análisis paramétricos.



Es una familia de estadísticos *univariados*:

- i) Informa sobre la distribución de una sola variable, por lo que no permite hacer inferencias sobre otras distribuciones.
- ii) Serán necesarios tantos estadísticos calculados como la cantidad de variables sobre las cuales se esté interesado en resumir su distribución.
- iii) Si el objetivo que se busca es construir una tipología empíricamente generada una población agrupada en distintos grupos de interés (centros educativos en públicos o privados; municipios en regiones; empresas en ramas de actividad; etc), una acumulación de estadísticos de tendencia central para caracterizar por ejemplo no agrega por sí misma ninguna información nueva, no constituye ninguna forma de análisis multivariado por más que se estén analizando varias variables. Básicamente no existe en este caso un análisis de distribuciones conjuntas.



Como todos los estadísticos el objetivo es presentar un resumen de una distribución. Su pertinencia debe ser analizada en función de cuál es la pregunta o las proposiciones que orientan el análisis.



Las medidas de tendencia central se diferencian por la complejidad requerida para su cálculo y por los supuestos que implican en su formulación. Los tres estadísticos más utilizados son:

- La moda o modo de la distribución, que se define como el valor que ocurre con más frecuencia (en términos absolutos).
- La mediana es un estadístico que informa en un conjunto de valores ordenados de menor al mayor, cuál es aquel que agrupa por lo menos a la mitad de las unidades observadas.

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

- La media aritmética o promedio se define como un valor típico de la distribución que tiende a ubicarse entre los valores intermedios observados.



Existen otras medidas de tendencia central de menor uso pero que resultan de interés en ciertos casos puntuales:

- La media aritmética ponderada, o promedio ponderado, que se caracteriza por ser un estadístico que asigna distintos pesos o ponderadores a los valores observados.
- La media geométrica
- La media armónica
- La media cuadrática
- Los cuartiles, deciles y percentiles

C. La moda o modo de una distribución de datos se define por el valor más frecuentemente observado.

- Puede ocurrir (y de hecho ocurre con frecuencia) que la distribución de una variable presente varias “modas”, es decir son varios los valores más frecuentes que se han registrado entre las unidades analizadas.
- Si la distribución tiene una única moda, se denomina *unimodal*. Si tiene dos modas, *bimodal*, así sucesivamente.
- Es una práctica recomendable graficar la distribución mediante un histograma o curva de frecuencias simples, para observar cuántas modas tiene la distribución y donde se ubican.

D. La **mediana** de una distribución es el valor que acumula la mitad de las unidades observadas, comenzando desde la que tiene el valor más bajo en la variable.



Otra forma de fijar la noción de mediana es la de “valor que está en el centro de una distribución”. El estadístico “parte” a las unidades observadas en dos mitades iguales e identifica el valor registrado del primer 50%+1 de las unidades como el “valor central”.

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

 La distribución debe estar ordenada para calcular la mediana:

- Para calcular la mediana de una distribución es necesario ordenar las unidades colocando en el primer renglón de la matriz de datos aquella que registra el valor más bajo en la variable de interés.
- Si se ordenan las unidades de mayor a menor se obtendrá el mismo valor de estadístico. Veáanse los ejemplos 1 y 2.

 Formalmente, para ubicar el número del caso (“*i*-ésimo caso”) en que se debe observar el valor de la Mediana (“Me”) de la variable de interés (“X”), se puede aplicar la siguiente fórmula:

$$X_i (\text{Me}) = (N + 1) / 2$$

 El valor de la mediana puede ser observado o calculado.

- Si el N total es un número *impar*, el valor será observado y se corresponderá con el valor de la variable que tiene la unidad ubicada inmediatamente después al de la mitad.
- Si el N total es un número *par*, el valor de la mediana será calculado y se corresponde con la semisuma entre los dos valores que ocupan las posiciones medias de la distribución ordenada.
- Ejemplo 1 (valor observado): el conjunto de nueve números ordenados de menor a mayor 2,2,3,4,6,7,8,8, y 10 tiene mediana 6.
- Ejemplo 2 (valor calculado): el conjunto de números ordenados 5,5,7,9,11,12,15 y 18 tiene mediana 10 y se calcula utilizando sólo los valores centrales:

$$\text{Me} = (9 + 11) / 2$$

 La mediana es un estadístico que tiene por propiedad no ser sensible a la presencia de valores extremos en la distribución.

- Ejemplo 3: En el conjunto de valores mostrados en el ejemplo 1 se han sustituido los dos primeros y los tres últimos de tal forma que tenemos ahora:

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

0,0,3,5,6,7,80,80, y 100. La mediana sigue siendo 6.

- Este propiedad puede resultar de interés para resumir una variable cuando la forma de la distribución de la variable está recargada hacia la izquierda, (con lo cual se dice que tiene “sesgo positivo”). Sin embargo, es conveniente mostrar la distribución mediante la gráfica respectiva.

E. La media aritmética o promedio es un estadístico que resume una distribución mediante un valor que “equilibra” todos los valores registrados.

 La media de un conjunto de N valores observados, $X = 1,2,3,4,\dots,N$, se denota por \bar{X} (X barra) y se obtiene:

$$\bar{X} = \frac{(X_1 + X_1 + X_2 + X_3 + \dots + X_N)}{N}$$

o

$$\bar{X} = \frac{\sum_{I=1}^N X_i}{N}$$

 De la expresión se pueden hacer las siguientes observaciones:

- Las X_i representan a cada uno de los valores válidos registrados. Es decir, se suman los N valores válidos
- El “modelo” está normalizado: esto permite comparar promedios originados en distintas poblaciones o en distintos tiempos.
- Cada uno de los valores tiene el mismo peso (o ponderación) en el cálculo de la media aritmética. Esta es una propiedad importante a tener presente cuando los valores que se resumen con este estadístico son porcentajes o proporciones atributos de un colectivo, como en el caso en que se promedian tasas de analfabetismo por municipios. Este promedio de porcentajes no es igual al porcentaje o proporción que se obtendría de calcular la tasa de

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

analfabetismo a partir de los datos individuales (es decir persona a persona para todo el universo). La razón es que todos los municipios pesan igual, sin importar cuál sea la población.

F. La media aritmética ponderada es un estadístico de tendencia central similar a la media aritmética simple pero en cuyo cálculo se supone que los valores sumados no tienen el mismo peso o ponderación.

- Formalmente:

$$\bar{X} = \frac{\sum (X_i * n_i)}{N}$$

o

$$\bar{X} = \sum (X_i * h_i)$$

- La expresión anterior indica que se puede calcular tanto multiplicando cada caso por su frecuencia absoluta dentro del agrupamiento, como multiplicando el caso por su proporción (obsérvese que en este caso no se divide el resultado entre N).
- Su utilización es apropiada para el cálculo de promedios en el caso de que los datos estén agrupados en "intervalos". Por ejemplo, en el caso de que la edad de los encuestados se agrupan por tramos de 5 años o en el caso de que el ingreso se haya preguntado por tramos de 500 pesos.
- Cuando se calcula un promedio para una variable propiedad analítica de un colectivo (por ejemplo, un municipio, una organización, un país) que ha sido construida promediando una variable individual (edad, ingresos, número de miembros del hogar, etc) deberá calcularse una gran media ponderando cada colectivo por el número de registros individuales .
- Estrictamente hablando, la media aritmética simple es un caso particular de la media aritmética ponderada en el cual el peso que tiene cada uno de los valores sumados es igual a 1

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

G. La media aritmética tiene un conjunto de propiedades de utilidad para su aplicación tanto en análisis concretos como para el desarrollo de otros estadísticos.

- Para la presentación y demostración de las propiedades de la media aritmética se utiliza la fórmula de la media ponderada.

 Propiedad 1: la suma de las desviaciones de los valores de la variable respecto de su media es igual a cero.

- Si se ha presentado este estadístico como una medida que se ubica en el “centro de gravedad” de la distribución, es razonable esperar que los valores que quedan a la izquierda y a la derecha “pesen” lo mismo.

$$\frac{\sum (x_i - \bar{x})n_i}{N} = \sum (x_i - \bar{x})h_i = 0$$

 Propiedad 2: el promedio de una constante es igual a la constante.

$$\bar{x}_k = k$$

 Propiedad 3 : al sumar una constante a una variable el promedio de la nueva distribución será igual al promedio original más la constante.

$$\bar{X}(K_1 + X_I) = K_1 + \bar{X}$$

 Propiedad 4: si se multiplican los valores de una variable por una constante, el promedio de la variable resultante es igual al promedio original multiplicado por la constante.

$$\bar{X}(K_1 * X_I) = K_1 * \bar{X}$$

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

-  Propiedad 5: si todos los valores de una variable se les multiplica por una constante K_1 y se le suma otra constante, K_2 , la nueva media aritmética será igual a multiplicar la media original por la constante K_1 y luego sumarle K_2 .

$$\overline{X}(K_2 + K_1 * X_I) = K_2 + K_1 * \overline{X}$$

-  Propiedad 6: el promedio general (o gran media) de un conjunto de observaciones agrupadas según otra variable, será igual al promedio de los promedios de cada categoría ponderados cada uno por el número de casos de cada categoría.

- ➡ Esta propiedad aplica la definición más arriba dada sobre la media aritmética ponderada.

H. Para concluir esta revisión, se presentan a continuación algunas características de las medidas de tendencia central que conviene tener presentes.

-  Estos estadísticos entregan valores que no necesariamente han sido observados para la matriz de datos. El promedio de número de hijos nacidos vivos por mujer en edad fértil constituye un indicador muy utilizado en demografía. Sin embargo, obsérvese que el valor 2,1 que este indicador toma para México, no existe empíricamente. Otro ejemplo análogo es el ingreso promedio de los hogares: la validez de este valor no depende de que exista empíricamente uno o varios hogares que tengan dicho ingreso.

-  En consecuencia, el significado sustantivo que se le da a una medida de tendencia central está mediado tanto por el nivel de medición como por las preguntas sustantivas que se buscan responder mediante ella. En muchos casos, dicho significado será necesariamente abstracto y bastante alejado del sentido común.

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

 La media aritmética es una excelente medida de tendencia central para los casos en que la distribución no presente sesgos importantes debido a la presencia de casos extremos.

- ✎ Supongamos que los ingresos traducidos a dólares de un conjunto de 10 hogares de la región "A" es tal como se lo presenta en el siguiente cuadro. Sobre esta base se calcula la media aritmética, la cual equivale a: 796 U\$S
- ✎ En la región "B" se observan prácticamente los mismos salarios, con la diferencia de 5 hogares donde el ingreso se ha multiplicado por 10. La media del ingreso de los hogares ahora asciende a 3101 U\$S
- ✎ En la región "C" en cambio se han modificado los mismos 5 hogares dividiendo el ingreso original entre 10. La media aritmética ahora calculada ha descendido a 532 U\$S.

 En consecuencia, si el propósito es informar cual es el valor predominante de una variable que resume su distribución en un conjunto de unidades, puede ser una respuesta que habrá que darse atendiendo a la forma de la distribución.

- ✎ Si la distribución está sesgada hacia alguno de los dos lados, probablemente convenga más utilizar la mediana. Tal fue la decisión que adoptara el Laboratorio Latinoamericano de Evaluación de la Calidad (LLECE) para informar los resultados de aprendizaje en Matemática para los alumnos de 4^o grado de Primaria en 13 países de América Latina.
- ✎ Si por el contrario la distribución no está sesgada probablemente sea más conveniente recurrir a la media aritmética, en razón de sus más conocidas propiedades y su mayor utilidad en otros análisis.

MEDIDAS DE TENDENCIA CENTRAL (Guía de clase)

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"
Hogar 1	828,2	828,2	828,2
Hogar 2	675,0	675,0	675,0
Hogar 3	603,3	603,3	603,3
Hogar 4	804,8	804,8	804,8
Hogar 5	400,0	400,0	400,0
Hogar 6	1066,7	1066,7	1066,7
Hogar 7	2145,0	2145,0	2145,0
Hogar 8	635,5	635,5	635,5
Hogar 9	760,5	760,5	760,5
Hogar 10	525,0	525,0	525,0
Hogar 11	1731,2	17312,0	173,1
Hogar 12	279,0	2790,0	27,9
Hogar 13	275,0	2750,0	27,5
Hogar 14	482,6	4820,6	48,3
Hogar 15	475,2	4750,2	47,5
Hogar 16	333,3	333,3	33,3
Hogar 17	525,0	525,0	52,5
Hogar 18	781,8	781,8	781,8
Hogar 19	992,3	992,3	992,3
Hogar 20	722,7	722,7	722,7
Hogar 21	625,0	625,0	625,0
Hogar 22	1274,3	12743,0	127,4
Hogar 23	1373,5	13735,0	137,4
Media aritmética	796,29	3101,08	532,20

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.