

## Escalas de medición estadística

### Statistical measurement scales

*Michel Enrique Gamboa Graus<sup>1</sup>*

#### Resumen

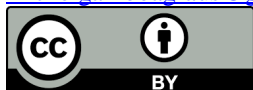
Este trabajo se enfocó en la utilidad de la Estadística para la investigación. Se presentaron varios aspectos relacionados con escalas de medición estadística, con énfasis en las escalas ordinales, como las más frecuentes en investigaciones educativas. Se presentó un procedimiento para ser utilizado en la elaboración de tesis en programas académicos de formación del profesional de la educación. Igualmente, se desplegaron aspectos esenciales para la comprensión de las escalas propuestas. Se introdujeron así criterios para valorar el nivel de ejecución de los indicadores y otorgar las categorías, la medición para cada miembro de la muestra y para la muestra como totalidad. Al mismo tiempo, se indicaron ideas para la organización y clasificación de los datos y categorías para evaluar las variables. Además, se mostraron ejemplos de cómo aplicar tales conceptos, en función del contexto y la finalidad de la investigación. Asimismo, se presentaron alternativas para la controvertida conversión de escalas ordinales a otras de intervalos y ejemplos de escalas que cumplen los requisitos establecidos. También se mostró cómo sintetizar los datos en valores representativos al llevar a cabo la indagación empírica, para tomar en cuenta el impacto de los resultados como totalidad.

*Palabras clave:* Estadística, escalas, medición

#### Abstract

---

<sup>1</sup> Licenciado en Educación, especialidades Matemática-Computación y Lenguas Extranjeras (Inglés). Doctor en Ciencias Pedagógicas. Centro de Estudios Pedagógicos de la Universidad de Las Tunas, Cuba. E-mail: [michelgamboagraus@gmail.com](mailto:michelgamboagraus@gmail.com). ORCID: <http://orcid.org/0000-0003-3704-9927>



This work focused on the usefulness of statistics for research. Several aspects related to scales of statistical measurement were presented, with emphasis in the ordinal scales, as the most frequent in educational research. A procedure to be used for the elaboration of thesis in academic programs of education professional training was presented. Likewise, essential aspects for the understanding of the proposed scales were deployed. Thus, criteria were introduced to evaluate the level of execution of the indicators and to grant the categories, the measurement for each member of the sample and for the sample as a whole. At the same time, ideas were indicated for the organization and classification of the data and categories to evaluate the variables. In addition, examples of how to apply such concepts were shown, depending on the context and purpose of the research. Also, alternatives for the controversial conversion of ordinal scales to interval scales and examples of scales that meet the requirements were presented. It was also shown how to synthesize the data into representative values when carrying out the empirical research, to take into account the impact of the results as a whole.

*Keywords:* Statistics, scales, measurement

### **Introducción**

Medir variables conduce a aprender sobre ellas y al compromiso de hacer algo al respecto. Esto lleva al constante crecimiento de nuestro entendimiento, desde una compleja concepción de la relación entre conocimiento y acción. Las continuas mediciones ofrecen valiosa información actualizada que permite desarrollar proyectos de investigación más pertinentes y satisfacer mejor las necesidades. Es eficiencia, en tanto las mediciones oportunas y acertadas optimizan la toma de decisiones. Además, es desarrollo, pues mejora lo que se mide. Medir, pues, se enmarca en una rica visión de la relación entre conocimiento y toma de decisiones.

Las definiciones más populares de medición en la literatura científica la presentan como la asignación de números a propiedades de objetos y eventos de acuerdo con reglas bien definidas (Stevens, 1946). Sin embargo, discrepo con esta arraigada posición reduccionista que excluye la posibilidad de representar propiedades mediante categorías cualitativas y hace asociar la medición a un proceso cuyo desenlace es irremediabilmente cuantitativo.

Así es que, para la investigación educativa, prefiero ver la medición como el proceso de vinculación de conceptos abstractos con indicadores empíricos (Zeller y Carmines 1980), que se realiza mediante un plan explícito y organizado para clasificar los datos disponibles. De tal forma, la medición se centra en la relación entre la respuesta observable y la construcción teórica, en función de lograr una válida y confiable correspondencia entre las propiedades y el sistema de categorías teóricamente construido. Cuando esta relación es fuerte, un análisis de los indicadores empíricos puede aportar información útil sobre los conceptos de interés.

El efecto de las escalas de medición estadística en la calidad de los datos procesados es uno de los temas de mayor relevancia para las investigaciones actuales. La idoneidad de un proceso de investigación se basa esencialmente en la calidad de estos datos en función de las necesidades del estudio. Sin embargo, es significativo el número de investigadores que no estudian los aspectos relacionados con la medición, lo que afecta el cumplimiento de sus objetivos generales y específicos. Al respecto, “la aplicación de un enfoque inadecuado de la medición en su estudio puede generar datos inapropiados” (Mendoza y Garza, 2009, p.18). De ahí la necesidad de desarrollar instrumentos de medición adecuados. Esto “implica aspectos como la precisión, la oportunidad, la accesibilidad, la interpretabilidad o la coherencia durante su uso y gestión de los mismos” (Matas, 2018, p.39).

También es necesario valorar la calidad metodológica, con el uso de medidas fiables y válidas, de técnicas analíticas apropiadas, los argumentos para la elección de categorías, entre otros aspectos. En correspondencia, Cascaes, Valdivia, da Rosa, Barbosa y da Silva (2013) muestran escalas que difieren entre sí en relación con el número de ítems, validez, fiabilidad y márgenes de puntuación. Esto evidencia que “la evaluación de la calidad de los estudios se torna esencial por su transparencia, visibilidad, rigor e impacto de la producción y publicación científica” (p. 307). Los aciertos o errores en el proceso de medición pueden afectar la validez de los datos, con lo que se puede perder la correspondencia entre el resultado de la medición y la realidad del fenómeno que se está estudiando (Sánchez y Echeverry, 2004). En tal sentido:

La medición es un proceso inherente y consustancial a toda investigación, sea ésta cualitativa o cuantitativa. Medimos principalmente variables y ello demanda considerar tres elementos básicos: el instrumento de medición, la escala de medición y el sistema de unidades de medición. La validez, consistencia y confiabilidad de los datos medidos dependen, en buena parte, de la escala de medición que se adopte. He ahí la importancia de profundizar en el tema de las escalas de medición. (Coronado, 2007, p.104)

De tal forma, existe una necesidad creciente de utilizar escalas de medición estadística en el proceso de investigación educativa para resolver, con mayor eficacia, los problemas de carácter investigativo de la práctica profesional. Como regularidad, “en los modelos teóricos se plantean un conjunto de relaciones entre variables no directamente observables y cuantificables de la realidad, pero que sí lo pueden ser a través de un grupo de variables manifiestas, ítems o indicadores” (Zapata y Canet, 2008, p.5). Sin embargo, en la realidad se presentan inconsistencias en la lógica consecuente del diseño de los modelos de investigación y la interpretación de sus resultados.

Una escala de medición estadística “es el conjunto de los posibles valores que una cierta variable puede tomar” (Coronado, 2007, p.106). En general, las escalas “son instrumentos de medición que comprenden un conjunto de ítems y estos ítems permiten identificar distintos niveles de las variables teóricas que no son directamente observables” (Mendoza y Garza, 2009, p.19). Son esenciales en las diferentes etapas de la investigación estadística. Espero que con este trabajo que presento se fortalezcan los conocimientos sobre el planeamiento o preparación, la recopilación de los datos, su organización y presentación, así como su síntesis, análisis e interpretación para la formulación de conclusiones y toma de decisiones.

Este trabajo se concentra en algunos de los aspectos que mayores dificultades he encontrado en las investigaciones que dirijo y evaluó regularmente. Al respecto, aquí presento temas como el de recoger información y clasificarla con escalas de medición, la interpretación de la misma en diferentes contextos, así como establecer tendencias de grupo de datos, realizar comparaciones y determinar relaciones entre indicadores para las proyecciones de variables a corto, mediano y largo plazo.

## **Desarrollo**

### **Materiales y métodos**

El criterio de expertos lo implementé para valorar la comprensibilidad, coherencia teórica y viabilidad de las escalas que propongo como ejemplos en el apartado de resultados. Así evalué su validez de apariencia. Al respecto, utilicé la propia autovaloración de los potenciales expertos con la aplicación del libro Excel CompetEx (Gamboa, 2019c). El libro Excel ComparEx (Gamboa, 2019b) lo utilicé para el procesamiento de los datos con la metodología de la comparación por pares. Con respecto a las categorías otorgadas por los expertos, de manera global los aspectos fueron evaluados de Bien (B) o Muy Bien (MB) para cada una de las escalas.

La validez de contenido o constructo la realicé con el análisis factorial exploratorio. Apliqué además el coeficiente de correlación de Pearson para cada uno de los casos, en función de evaluar cómo se relacionan los puntajes de los diferentes momentos, en términos de asociación lineal. Igualmente, utilicé análisis de varianza para mediciones repetidas en función de valorar la sensibilidad al cambio. A la par, la medición de la consistencia interna u homogeneidad de cada escala de los ejemplos la realicé mediante el coeficiente alfa de Cronbach.

Con respecto a la determinación de la utilidad de las escalas, en función de que fueran fácil de aplicar y procesar como referencia a su aplicabilidad en escenarios auténticos, estas fueron utilizadas en investigaciones reales. De tal forma valoré tanto el tiempo promedio requerido para aplicarlas, el grado de capacitación que requieren quienes se debían encargar de hacerlo, así como los procedimientos y tiempos requeridos para calificar sus puntajes. Al respecto, en ocasiones se hizo muy complicado el procesamiento de números tan elevados de datos, junto a la presentación distorsionada de los valores obtenidos, con la omisión de algunos valores claves que conducía a una valoración incompleta o parcial de los resultados.

Lo anterior, más que una dificultad se convirtió en una oportunidad por aprovechar. Así es que elaboré los respectivos productos informáticos de cada una de las escalas como herramientas para extraer información accionable de la gran cantidad de datos disponibles. Estos se emplean para la organización y clasificación de la información recopilada, de modo que se facilite la presentación automática en tablas y gráficos en función de reflexionar sobre el comportamiento de los datos. Así, entre sus principales ventajas puedo referir que permiten disponer de mayor cantidad de información, obtener respuestas más completas que reflejan mayor fiabilidad de datos y un enfoque más preciso para abordar problemáticas. La adecuada

utilización y explotación de estos datos constituye una gran oportunidad. Las oportunidades se multiplican a medida que se aprovechan. Este escenario contribuyó a la reproducibilidad del proceso investigativo. De tal forma se hicieron públicos, no solo los informes de investigación, sino también las bases de datos conformadas y los algoritmos empleados para procesarlos estadísticamente.

### **Resultados y discusión**

Medir es un proceso inherente a las investigaciones científicas. En tal sentido, “la medición es una actividad fundamental, que busca que el proceso de observación de personas, objetos, entre otros aspectos de la realidad, tenga sentido” (Mendoza y Garza, 2009, p.18). Para medir las variables que se estudian, además del instrumento de medición, la escala es esencial. Por ejemplo, el procedimiento que recomiendo para utilizar en la elaboración de tesis en programas académicos de formación del profesional de la educación incluye los siguientes pasos:

- Planeamiento o preparación (Conceptualización y operacionalización de la variable. Descripción de la escala. Selección, o elaboración y validación del instrumental científico. Validación de la escala. Selección de la muestra de investigación. Preparación del equipo de aplicadores del instrumental científico).
- Recopilación de los datos (Aplicación del instrumental científico).
- Organización y presentación de datos (Procesamiento de la información. Tabulación, organización y presentación en tablas, cuadros o gráficos).
- Síntesis, análisis e interpretación de los datos (Caracterización del nivel de desarrollo de la variable).
- Formulación de conclusiones y toma de decisiones (Formulación de conclusiones y toma de decisiones).

El proceso de recolección de información debe dirigirse fundamentalmente a la variable dependiente. Los instrumentos deben enfocarse a los elementos que caracterizan a dicha variable, a sus dimensiones e indicadores, o a los elementos de las otras variables que están presentes en el proceso, y tienen una estrecha relación con ella (causas). En la figura muestro una caricatura, que suelo presentar con regularidad en mis clases y que mi hija no para de reír cuando la ve, de cómo sería la tesis de un investigador que no considere adecuadamente el proceso de medición para arribar a conclusiones válidas.

Las escalas ordinales son las más frecuentes en investigaciones educativas, donde suelen abundar variables cualitativas (Gamboa, 2018). Nos otorga la clasificación y el orden de los datos sin que realmente se establezca el grado de variación entre ellos. Por su forma de representación pueden ser numéricas, verbales o gráficas. Al respecto, “Para efectuar la medición en estos casos se recurre a una estrategia que es agrupar las características de la entidad en categorías” (Sánchez y Echeverry, 2004, p.304). Estas, pueden aparecer en la literatura sobre el tema como constructos o variables latentes, son “aspectos observables o características propias del mundo empírico (...) no medimos objetos, personas o colectividades en cuanto tales; medimos propiedades observables de ellas” (Coronado, 2007, p.104-105). Se utiliza una escala ordinal para comprender si estas propiedades son mayores o menores, mejores o peores.

Una escala ordinal, además de identificar y describir la magnitud de la variable, suele mostrar su rango relativo. Quizás la ventaja principal de utilizarla es la facilidad de comparación que ofrece. Es la ideal para agrupar después de ordenar. Al respecto, la identificación y definición de indicadores de las variables que van a ser medidas son esenciales.

Es preciso explicitar adecuadamente las categorías de las escalas pues existen sesgos asociados a estilos de respuesta (Paulhus, 1991). Entre ellos destacan la tendencia a elegir las



opciones extremas o intermedias, independientemente de su contenido (He y Von De Vijver, 2015), conjuntamente con la que sea mejor vista por los demás, como deseabilidad social (Nadler, Weston y Voyles, 2015).

Otro aspecto crucial por considerar es que el número de alternativas afecta a las propiedades psicométricas de las escalas. Churchill y Peter (1984) presentaron una correlación positiva entre la fiabilidad y el número de puntos de los ítems de la escala, de forma que aumentar el número de opciones aumenta la fiabilidad. Al respecto, “La mayoría de investigaciones utilizan entre 3 y 7 puntos en las escalas” (Bisquerra y Pérez, 2015, p.132). Sin embargo, “Un argumento psicométrico para justificar escalas de 5-7 puntos es que diversos estudios observaron que la presencia de más puntos en los ítems no aumenta la fiabilidad de la prueba” (Bisquerra y Pérez, 2015, p.130), de manera que la confiabilidad aumenta cuando las alternativas se incrementan de cinco a siete (Dawes, 2008), siendo menos apreciable a partir de siete (Dillman, 2007). Igualmente, la alternativa intermedia está estrechamente vinculada a los sesgos de tendencia central y deseabilidad social (Baka & Figgou, 2012), si bien las personas la seleccionan por diversos motivos, independientemente de su real posicionamiento (Velez y Ashworth, 2007).

Al respecto, he realizado investigaciones que van desde solo dos alternativas (Gamboa, 2007), hasta 11 posibilidades (Gamboa y Borrero, 2016). He querido tener experiencias en investigaciones reales con diferentes números de opciones, para valorar por mí mismo sus pros y contras. Como consecuencia, para los estudios que realizo prefiero la posibilidad de seis opciones, con aspectos de exploración específicos en un escalamiento categórico, de forma que la escala sea práctica, fácil de aplicar y de procesar. Esto incrementa su utilidad para otros estudios, por otros investigadores. Como regularidad, busco que se pueda aplicar en un tiempo

razonable para la investigación, que no requiera demasiado entrenamiento para potenciales aplicadores, y que sea sencillo de calificar.

Ir de los conceptos definidos teóricamente a los definidos operativamente no es directo. Asimismo, comprender y seleccionar el modelo estadístico apropiado para decidir qué datos recoger y valorar, para emplearlo en una investigación determinada, pasa por un cúmulo de decisiones importantes que difícilmente se toman de inmediato. La cantidad y la calidad de los datos recolectados es esencial para las investigaciones educativas. Sin embargo, “contra lo que a menudo se piensa, el proceso de desarrollo de estas herramientas es necesariamente largo, si se quiere que sea sólido” (Martínez, 2010, p.14), y no está exento de errores. El tiempo que implica no sólo el primer planteamiento, sino su refinamiento hasta alcanzar un grado aceptable de madurez, es dilatado. A mí me ha pasado que he tardado años de ir y venir perfeccionando una escala que elaboré inicialmente para un estudio particular, en función de implementarla en estudios posteriores. En consecuencia, la primera recomendación que le doy a mis estudiantes de programas académicos es revisar la literatura para ver si existen escalas que se ajustan a sus propósitos, y así tomar la decisión de construir una nueva o no, para evitar errores innecesarios.

### **Aspectos esenciales para la comprensión de las escalas que propongo**

En este apartado presento aspectos que son esenciales para la comprensión de las escalas que propongo. Como parte del planeamiento o preparación para usarlas destaco que están compuestas por características construidas según escalas de tipo Likert, de seis alternativas. La idea es que proporcionen puntuaciones graduadas a una serie de enunciados.

Presento escalas ordinales en las que cada indicador muestra una característica en el objeto de investigación y se mide con una escala de 6 puntos de recorrido. Esto lo significo con categorías que, en términos cualitativos, evalúan el nivel de ejecución de dicha característica en

el proceso. Los criterios para valorar el nivel de ejecución de cada uno de estos indicadores, y otorgar las categorías correspondientes, son los siguientes:

- Nulo, (N-0): No se muestran evidencias de la presencia del indicador en el proceso. Luego, el nivel de ejecución del indicador es inexistente. Esta es la categoría inferior porque revela la necesidad de lograr que el conocimiento emerja de la ignorancia.
- Mal, (M-1): Se muestran evidencias de la presencia del indicador en el proceso. Sin embargo, la ejecución es contraria a lo que es debido, desacertadamente, de mala manera. Contrariamente a lo que se requiere, de manera impropia o inadecuada para el logro de la formación integral. De tal forma, es nociva y daña esta aspiración. Este criterio recibe una valoración superior a la no presencia del indicador a partir de que los errores “son una utilísima semilla que hay que saber tratar para que llegue a flor y a fruto” (Gamboa, 2020b, p.54). Al error no debe temérsele, es uno de los mejores pretextos para mejorar. Es preferible que se tenga la conciencia para hacerlo, aunque se necesite ayuda para hacerlo bien. Esta categoría revela la necesidad de lograr que el conocimiento emerja de uno previo, aunque esté equivocado.
- Regular, (R-2): Se muestran evidencias de la presencia del indicador en el proceso. No obstante, la ejecución es parcialmente correcta, no bien. De mediana calidad o intermedia. Que se encuentra en un término medio entre las categorías de Mal y Bien. Que contiene algún error de menor importancia. Igualmente, si tiene presencia el indicador, pero no se desarrolla lo suficiente para tener un impacto apreciable. Esta categoría revela la necesidad de investigar de qué manera el conocimiento incompleto e inexacto llega a ser más completo y más exacto.

- Bien, (B-3): Se muestran evidencias de la utilidad y el beneficio de la presencia del indicador en el proceso. Se demuestra que la ejecución posee un valor positivo y por ello es estimable. Se realiza según es debido, con razón, acertadamente, de buena manera, de buen proceder. Sin inconveniente o dificultad. Según se requiere, de manera propia o adecuada para el logro de la formación integral. De tal forma, se estima conveniente, de conformidad con esta aspiración. Sin contradicción antagónica.
- Muy Bien, (MB-4): Se muestran evidencias de que la presencia del indicador en el proceso es extraordinariamente buena, que sobresale en méritos. Se demuestra que la ejecución sobresale por sus óptimas cualidades, que es muy buena o que sobresale en alguna cualidad con respecto a la exigencia ordinaria, aunque sin crear algún producto con el que defienda su punto de vista, justifique su postura y muestre conexiones según la organización del tema. El nivel de ejecución sobresale en cuanto a su calidad, superioridad, y por ello es que es objeto de una estima y de una valoración elevada, y además sobresale de la media.
- Excelente, (E-5): Se muestran evidencias de que la presencia del indicador en el proceso es extraordinariamente buena. Además, se demuestra profundidad y un nivel creativo de ejecución del indicador, con algún producto personal en el que defienda su punto de vista, justifique su postura y muestre conexiones según la organización del tema, mostrando madurez en el uso o manejo del mismo.

“Los indicadores son considerados como las unidades básicas de valoración. Las categorías determinan el peso que cada criterio recibe al valorar el indicador” (Gamboa, 2020c, p.150). Cada criterio señala los aspectos por considerar dentro de la evaluación del indicador. Si el análisis se hará a nivel ordinal se toman las letras (N, M, R, B, MB, E). Si el análisis se hará a

nivel intervalar se toman los números (0, 1, 2, 3, 4, 5). Más adelante mostraré ejemplos. También es importante recordar que el uso de estas escalas “se hace en un proceso de valoración auténtica, situados en el proceso mismo” (Gamboa, 2020c, p.150).

Se pueden utilizar variantes de esta propuesta para cinco puntos de recorrido (N, M, R, B, MB). También se puede adaptar para cuatro puntos de recorrido (M, R, B, MB). Un ejemplo sería como sigue:

- Mal (M): No se muestran evidencias de la presencia del indicador en el proceso. Luego, el nivel de ejecución del indicador es inexistente. Igualmente, si se muestran evidencias de la presencia del indicador en el proceso, pero la ejecución es contraria a lo que es debido, desacertadamente, de mala manera. Contrariamente a lo que se requiere, de manera impropia o inadecuada para el logro de la adecuada dirección educacional. De tal forma, es nociva y daña esta aspiración. Esta categoría revela la necesidad de lograr que el conocimiento emerja de la ignorancia o de uno previo equivocado.
- Regular (R): Se muestran evidencias de la presencia del indicador en el proceso. No obstante, la ejecución es parcialmente correcta, no bien. De mediana calidad o intermedia. Que se encuentra en un término medio entre las categorías de Mal y Bien. Que contiene algún error de menor importancia. Igualmente, si tiene presencia el indicador, pero no se desarrolla lo suficiente para tener un impacto apreciable. Esta categoría revela la necesidad de investigar de qué manera el conocimiento incompleto e inexacto llega a ser más completo y más exacto.
- Bien (B): Se muestran evidencias de la utilidad y el beneficio de la presencia del indicador en el proceso. Se demuestra que la ejecución posee un valor positivo y por ello es estimable. Se realiza según es debido, con razón, acertadamente, de buena manera, de

buen proceder. Sin inconveniente o dificultad. Según se requiere, de manera propia o correcta para el logro de la adecuada dirección educacional. De tal forma, se estima conveniente, de conformidad con esta aspiración. Sin contradicción antagónica.

- Muy Bien (MB): Se muestran evidencias de que la presencia del indicador en el proceso es extraordinariamente buena, que sobresale en méritos. Se demuestra que la ejecución sobresale por sus óptimas cualidades, que es muy buena o que sobresale en alguna cualidad con respecto a la exigencia ordinaria. El nivel de ejecución sobresale en cuanto a su calidad, superioridad, y por ello es que es objeto de una estima y de una valoración elevada, y además sobresale de la media.

Las propuestas que hago comienzan siempre con la medición para cada miembro de la muestra. Busco que cada indicador o sub-indicador se mida desde diferentes perspectivas. En este sentido, se debe evaluar con la aplicación de diversos métodos, técnicas e instrumentos ( $M_n$ ) para contrastar los resultados. En correspondencia, se deben buscar los promedios per cápita (mediana o media aritmética según sea el caso de muestra pequeña o grande) de las puntuaciones obtenidas en la escala Likert utilizada. Como regularidad, se debe tener cierto equilibrio en la cantidad de métodos, técnicas o instrumentos utilizados para medir cada indicador o sub-indicador. Así se evitarían errores, al impedir que influya el peso que tendrían grandes diferencias en la valoración global. Luego, estos son utilizados en función de obtener la evaluación de cada una de las dimensiones, así como la variable para cada uno de los muestreados.

La medición de la variable para la muestra como totalidad se debe atender posteriormente. Esta se debe calcular utilizando los promedios calculados previamente para cada uno de los miembros de la muestra ( $N_n$ ). A este tenor, el promedio general de cada uno de los

indicadores o sub-indicadores para la muestra es el que brinda las conclusiones generales de la variable. Así se pondera el peso que tendría cada uno de los muestreados. Vale la pena invertir tiempo en el análisis de los datos. Nunca se escala en vano en las montañas de la verdad.

Después de la recopilación de los datos, se debe proceder a su organización y clasificación. Recomiendo utilizar los libros Excel que elaboré para cada una de las escalas que propongo, de modo que se facilite la presentación en tablas y gráficos. Aquí enfatizo en la necesidad de realizar una evaluación y ajuste de los datos, con el propósito de superar las omisiones o errores. La validez de los resultados y conclusiones dependen en gran medida de la fidelidad de los datos utilizados. En Gamboa (2019a) muestro el trabajo con las diferentes hojas de uno de estos libros Excel. Ahí presento el trabajo con una hoja para cada muestreado, tabulación general, gráficos, la selección de la hoja de trabajo, la introducción de datos para medir cada miembro de la muestra y la muestra como totalidad, así como el copiado de una hoja para cada muestreado para conservar evidencias y las condiciones para introducir los datos.

El procedimiento para las escalas que propongo es general, en función de la síntesis de los datos. Muestro algunos ejemplos a continuación. En cada uno de los casos, las categorías que se emplean para evaluar las variables se escogen según las medidas de tendencia central utilizadas, en una gradación desde la excelencia hasta niveles inferiores. Recomiendo la mediana para los casos que emplean muestras pequeñas, lo que brindaría directamente las categorías. Entretanto, la media aritmética sería una mejor opción para muestras mayores. La Tabla 1 muestra distintas variantes para otorgar las categorías si se empleara la media aritmética, dependiendo del número de estas. Estas escalas se utilizan tanto en la indagación empírica, como en la validación de las propuestas.

**Tabla 1***Variantes para otorgar categorías si se empleara la media aritmética*

Variantes	Posibles categorías					
	N	M	R	B	MB	E
<b>6 categ.</b>	$\bar{x} \leq 0,75$	$0,75 < \bar{x} \leq 1,75$	$1,75 < \bar{x} \leq 2,75$	$2,75 < \bar{x} \leq 3,75$	$3,75 < \bar{x} \leq 4,75$	$4,75 < \bar{x} \leq 5$
<b>5 categ.</b>	$\bar{x} \leq 0,75$	$0,75 < \bar{x} \leq 1,75$	$1,75 < \bar{x} \leq 2,75$	$2,75 < \bar{x} \leq 3,75$	$3,75 < \bar{x} \leq 4$	
<b>4 categ.</b>		$\bar{x} \leq 0,75$	$0,75 < \bar{x} \leq 1,75$	$1,75 < \bar{x} \leq 2,75$	$2,75 < \bar{x} \leq 3$	

*Fuente:* Elaboración propia**Ejemplos de escalas para valorar variables**

En estadística “el tipo de análisis depende del nivel o escala de medición de las variables de investigación (...) mientras más complejo o alto es el nivel de medición, más efectivos son los métodos estadísticos que se pueden utilizar” (Coronado, 2007, p.121). Los procedimientos estadísticos para el nivel intervalar proporcionan más información y con mayor detalle que el nivel ordinal. No obstante, el análisis de los cálculos no puede considerarlos como un fin, sino en función de una interpretación del comportamiento de los datos.

En este trabajo persigo dejar clara la idea de que “la medición no es un fin en sí misma, y sólo tiene legítimo sentido cuando se la percibe sirviendo a los fines instrumentales del conocimiento teórico y pragmático” (Coronado, 2007, p.105). Aunque categorizar la variable y las conclusiones cuantitativas son importantes, lo fundamental es la interpretación cualitativa que se debe hacer. En tal sentido, “el número para lo cualitativo no se separa de un proceso de significación que siempre lo trasciende (...) no es la respuesta puntual que vamos obteniendo sino los juegos que somos capaces de construir con la información de las escalas” (Díaz, González y Arias, 2017, p.137).

En correspondencia, las escalas que aquí presento tienen el propósito de generar grandes volúmenes de datos relevantes obtenidos de diferentes fuentes, procedentes de las diferentes



acciones de los evaluados. Al respecto, en Gamboa (2021a) se pueden encontrar múltiples ejemplos validados para la investigación educativa en contextos reales. Los datos son un tesoro muy atractivo. La intención es que se haga una recopilación de manera que permita encontrar patrones interesantes de información útil y novedosa, en función de descubrir relaciones y regularidades existentes, y sus posibilidades de generalización. Esto, más que para dar un valor de medición de las variables, es sobre todo para la búsqueda de causas en los patrones de comportamiento de los datos. El proceso que propongo va más allá de calcular o deducir, implica la observación de patrones, la comprobación de conjeturas y la estimación de resultados.

### ***Ejemplos de escalas para valorar variables a nivel ordinal***

La mediana es la medida de tendencia central más representativa para la valoración de variables a nivel ordinal. Esta es aplicable a cualquier tipo de datos que puedan ser ordenados, siempre existe y es única, y es ideal para las escalas ordinales como las que presento en este apartado. Es un valor de la variable que, una vez que están ordenados los datos de menor a mayor o viceversa, deja por debajo de sí a la mitad de los datos, y la otra mitad por encima. Se sitúa, por tanto, en la mitad real de los datos. Es el valor que equidista de los extremos. Es el punto medio del conjunto de datos. No es una función algebraica de los datos individuales, por lo que no está afectada por cada dato, ni principalmente por los valores extremos. De ahí que sea de las medidas estadísticas más robustas, apropiada para un grupo pequeño de datos. Es por eso que se asocia, fundamentalmente en estudios de muestras pequeñas, porque la mediana es una métrica de punto medio mejor para los casos en los que un número pequeño de valores atípicos podrían sesgar drásticamente la media. Sin embargo, también recomiendo utilizarla cuando los datos no se ajustan a una distribución normal.

Un ejemplo para valorar variables a nivel ordinal se puede verificar con el estudio de la

participación familiar en una educación escolar inclusiva (Gamboa y Hernández, 2016). En él se consideró la valoración de tres dimensiones, con 37 indicadores. En este caso se evaluaron con la escala ordinal (N, M, R, B, MB, E). Luego se buscó la mediana de las evaluaciones por indicador, dimensión y el total general para evaluar la variable. Esto es así, tanto para valorar la participación familiar en cada elemento de la muestra como para la muestra general. La mediana es apropiada para un grupo pequeño de datos como los que se aspira en esta escala para el trabajo con familias de una institución educativa. PaFEeI (Gamboa, 2016) es un software que permite aplicar la escala para valorar la participación familiar en una educación inclusiva, a partir de datos introducidos por el usuario. Otro ejemplo se puede encontrar con la escala estadística para evaluar coherencia didáctica en procesos de enseñanza-aprendizaje de Matemáticas, descrita en Gamboa (2021a). Para ello se puede utilizar el libro Excel EsCoDi ordinal (Gamboa, 2021c).

### ***Ejemplos de escalas para valorar variables a nivel intervalar***

El supuesto de continuidad se comparte en algunos casos, por su larga amplitud (Gamboa, 2018). La valoración de variables a nivel intervalar se realiza con la media aritmética como medida de tendencia central en estos casos. La media aritmética se aplica cuando la variable está medida en escalas métricas. Siempre existe, es única y fácil de calcular. Además, es una función algebraica de los datos individuales, por lo que está afectada por cada dato y principalmente por los valores extremos. Se desvía hacia estos valores, tanto más cuanto más extremos son. De ahí que recomiendo utilizarla solo para muestras grandes.

El cálculo de la media aritmética requiere una escala de intervalo. Algunos autores como DiStefano (2002) han argumentado que no es legítimo el uso de operaciones estadísticas propias de variables cuantitativas en presencia de datos ordinales. Sin embargo, Bollen & Barb (1981), Solís (2014), Asún, Rdz-Navarro y Alvarado (2016) y otros defienden la posibilidad y utilidad de

emplear estadística paramétrica para analizar datos ordinales en determinadas situaciones. Ellos han argumentado que es una alternativa aceptable, lo que también compartí en varios trabajos precedentes como Gamboa (2017), Gamboa y Parra (2017; 2019), Gamboa y Hernández (2018), Rodríguez, Gamboa y Oliva (2019). Considero que es legítimo analizar datos ordinales paramétricamente, por su utilidad para conducir a fructíferos resultados, como práctica que cuenta con el respaldo de la comunidad académica. Eso sí, hay que cumplir con los requisitos para ello (Rositas, 2014) y se debe trabajar con muestras mayores para conseguir una potencia equiparable a los análisis de datos numéricos.

La valoración de variables a nivel intervalar se realiza con la media aritmética como medida de tendencia central. La valoración de cuatro dimensiones, con cuatro indicadores per cápita, fue la esencia que consideramos para categorizar la Resolución de Problemas en el Proceso de Enseñanza Aprendizaje de la Matemática (Rodríguez y Gamboa, 2019). De tal forma, se buscan los promedios de las evaluaciones obtenidas en cada uno de los indicadores, los que son utilizados en función de obtener la resolución de problemas para cada miembro de la muestra. Para medir el comportamiento de las dimensiones e indicadores se elaboró el libro Excel ResProMa (Rodríguez y Gamboa, 2019), el cual ofrece una escala valorativa con sus correspondientes categorías.

La escala estadística para evaluar coherencia didáctica en procesos de enseñanza-aprendizaje de Matemáticas (Gamboa, 2020c) analiza también a nivel intervalar. Para ello se puede utilizar el libro Excel EsCoDi (Gamboa, 2019d). Igualmente se puede acceder a otros ejemplos que llegan incluso a la utilización de sub-indicadores, como la valoración de la competencia de dirección en educación (Gamboa, Castillo y Parra, 2019; 2020) con el libro Excel EsComDE (2019a) y la valoración del modo de actuación profesional médico (Mora,

Santiesteban y Gamboa, 2020) con el libro Excel MAPMe (Mora y Gamboa, 2019). Otro ejemplo en el que dejó claro que las dimensiones, indicadores o indicadores no necesariamente deben tener el mismo peso en los cálculos se puede encontrar en la tesis de pregrado de mi segunda carrera universitaria para mi especialización en lenguas extranjeras, en función de valorar el desempeño de los docentes para enseñar ciencias naturales y exactas en inglés (Gamboa, 2020a), así como en Gamboa (2021b), donde valoré la formación de profesores de matemáticas de habla hispana para enseñar en países de habla inglesa.

La esencia de los ejemplos que mostré hasta aquí está en el énfasis que quiero dar en evitar conclusiones sobre la base de datos que no han ocurrido en diferentes momentos. Por eso la idea de múltiples métodos, técnicas e instrumentos para evaluar cada indicador o sub-indicador. Esto previene a los investigadores de una toma de decisiones sobre la base de hechos aislados e insuficientes datos.

### **Conclusiones**

El mensaje fundamental de este artículo radica en que cada elemento de la investigación educativa, y en particular los relacionados con los estadísticos, no debe ser arbitrario. Estos deben ser planeados cuidadosamente para poder asegurar la validez y fiabilidad de los resultados que se obtengan, pues estos serán utilizados en la resolución de problemas externos a la propia Estadística. Las escalas de medición y el análisis de los datos deben estar regidos por el criterio de idoneidad para el estudio que se realiza. Las opciones sobre las que tomar decisiones deben considerarse conscientes del propósito de la investigación, el diseño de la misma, el tiempo y demás recursos con los que se cuenta, las restricciones del proceso, los métodos de colección de los datos, la metodología que se emplee. Las decisiones que se tomen para aplicar la Estadística a la investigación educativa deben ser en correspondencia con el contexto de la investigación. De

tal forma se favorece la adecuada recolección, organización, presentación y análisis de datos relativos a las muestras o poblaciones de estudio, para arribar a conclusiones válidas referidas a las variables que se miden y tomar decisiones razonables.

### Referencias

- Asún, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2016). Developing multidimensional Likert scales using item factor analysis: The case of four-point items. *Sociological Methods and Research*, 45(1), 109-133.
- Baka, A. & Figgou, L. (2012). 'Neither agree, nor disagree': a critical analysis of the middle answer category in voting advice applications. *International Journal Governance*, 5(3/4), 244-263.
- Bisquerra, R. y Pérez, N. (2015). ¿Pueden las escalas Likert aumentar en sensibilidad? *REIRE, Revista d'Innovació i Recerca en Educació*, 8 (2), 129-147.
- Bollen, K.A. & Barb, K.H. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review*, 46(2), 232-239.
- Cascaes, F., Valdivia, B. A., da Rosa, R., Barbosa, P. J. y da Silva, R. (2013). Escalas y listas de evaluación de la calidad de estudios científicos. *Revista Cubana de Información en Ciencias de la Salud (ACIMED)*, 24(3), 295-312.
- Churchill, G., y Peter, J. P. (1984). Research design effects on the reliability of rating scales: A Meta-Analysis. *Journal of Marketing Research*, 21(4), 360-375.
- Coronado, J. (2007). Escalas de medición. *Paradigmas*, 2(2), 104-125.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research*, 50(1), 1-19.

- Díaz, A., González, F. y Arias, A. M. (2017). Pensar el método en los procesos de investigación en subjetividad. *Rev. CES Psicol.*, 10(1), 129-145.
- Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method*. Nueva York: John Wiley y Sons.
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(3), 327-346.
- Gamboa, M.E. (2007). El diseño de unidades didácticas contextualizadas para la enseñanza de la Matemática en la Educación Secundaria Básica. Tesis Doctoral. Instituto Superior Pedagógico José de la Luz y Caballero.
- Gamboa, M.E. (2016). Libro Excel para aplicar la escala de participación familiar en una educación escolar inclusiva (PaFEI). <http://roa.ult.edu.cu/jspui/handle/123456789/4519>
- Gamboa, M.E. (2017). Estadística aplicada a la investigación científica. En J.C. Arboleda. (Ed.). *Apropiación, generación y uso solidario del conocimiento* (pp. 59-76). Las Tunas, Cuba: Editorial Redipe-Edacun.
- Gamboa, M.E. (2018). Estadística aplicada a la investigación educativa. *Dilemas Contemporáneos: Educación, Política y Valores*, 5(2).
- Gamboa, M.E. (2019a). Libro Excel EsComDE como recurso para medir la Competencia de Dirección en Educación. *Boletín Redipe*, 8(3), 149-184.
- Gamboa, M.E. (2019b). Libro Excel de comparación por pares para procesar criterios de expertos (ComParEx). <http://roa.ult.edu.cu/jspui/handle/123456789/3957>
- Gamboa, M.E. (2019c). Libro Excel para calcular la competencia de expertos (CompetEx). <http://roa.ult.edu.cu/jspui/handle/123456789/3958>

Gamboa, M.E. (2019d). Libro Excel para aplicar la escala de la coherencia didáctica (EsCoDi).

<http://roa.ult.edu.cu/jspui/handle/123456789/3955>

Gamboa, M.E. (2020a). Alternative to train Spanish speaking Math teachers for teaching in English speaking countries. *Didasc@lia: Didáctica y Educación*, 11(3), 106-131.

Gamboa, M.E. (2020b). Errores en el aprendizaje. Utilísima semilla que debe llegar a flor y a fruto. Editorial Académica Española.

Gamboa, M.E. (2020c). Escala estadística y software para evaluar coherencia didáctica en procesos de enseñanza-aprendizaje de Matemáticas. *Didasc@lia: Didáctica y Educación*, 11(1), 140-165.

Gamboa, M.E. (2021a). *Escalas de Medición Estadística. Ejemplos validados para la investigación educativa*. Editorial Académica Española.

Gamboa, M. E. (2021b). Formación de profesores de matemáticas de habla hispana para enseñar en países de habla inglesa. *Revista ROCA*, 17(4), 453-471.

Gamboa, M. E. (2021c). Libro Excel para aplicar la escala de la coherencia didáctica a nivel ordinal (EsCoDi ordinal). <http://roa.ult.edu.cu/jspui/handle/123456789/4529>

Gamboa, M.E. & Borrero, R.Y. (2016). Influencia de la contextualización didáctica en la coherencia curricular del proceso. *Revista Dilemas Contemporáneos: Educación, Política y Valores*. 4(1).

Gamboa, M.E., Castillo, Y. y Parra, J.F. (2019). Caracterización de la competencia de dirección en educación para el ejercicio pedagógico en el escenario educativo tunero. *Dilemas Contemporáneos: Educación, Política y Valores*, 6(3).

- Gamboa, M.E. y Hernández, L.J. (2016). Escala estadística y software para valorar la participación familiar en una educación escolar inclusiva. *Revista Mikarimin*, 2(3), 45-52.
- Gamboa, M.E. y Hernández, L.J. (2018). Escala para medir la educación inclusiva de la familia. En J.C. Arboleda (Ed.), *Educación y Pedagogía Cuba 2018* (2), (pp. 6-15). La Habana, Cuba: Editorial Redipe.
- Gamboa, M.E. y Parra, J.F. (2017). Diseño de una escala para medir la competencia de dirección en Educación. En E. Santiesteban y J. C. Arboleda (Eds.), *Ciencia e Innovación Tecnológica* (1), (pp. 542-552). Las Tunas, Cuba: Sello Editorial Edacun-Redipe.
- Gamboa, M.E. y Parra, J.F. (2019). *Recursos para investigar sobre Competencia de Dirección en Educación. Ejemplos de buenas prácticas en su aplicación*. OmniScriptum Publishing Group, Mauritius: Editorial Académica Española.
- He, J. & Von De Vijver, J. R. (2015). Effects of a general response style on cross-cultural comparisons. Evidence from the teaching and learning international survey. *Public Opinion Quarterly*, 79, 267-290.
- Martínez, F. (2010). Los indicadores como herramientas para la evaluación de la calidad de los sistemas educativos. *Sinéctica*, (35), 1-17.
- Matas, A. (2018). Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Revista electrónica de investigación educativa*, 20(1), 38-47.
- Mendoza, J. y Garza, J. B. (2009). La medición en el proceso de investigación científica: Evaluación de validez de contenido y confiabilidad. *Innovaciones de negocios*, 6(11), 17-32.



- Mora, K.M. y Gamboa, M.E. (2019). Libro Excel para investigar sobre el modo de actuación profesional médico en formación inicial (MAPMe). <http://roa.ult.edu.cu/jspui/handle/123456789/3967>
- Mora, K.M., Santiesteban, Y. y Gamboa, M.E. (2020). El modo de actuación profesional médico del estudiante de medicina en formación inicial. *Didasc@lia: Didáctica y Educación*, 11(6), 271-295.
- Nadler, J., Weston, R. y Voyles, E. (2015). Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, 142(2), 71-89.
- Paulhus, D. (1991). Measurement and control of response biases. En J. Robinson, P. Shaver y L. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-51). EUA: Academic Press.
- Rodríguez, L.A. y Gamboa, M.E. (2019). Ejercicios para favorecer la comprensión de problemas matemáticos en la educación de adultos. *Revista Cognosis*, 4(3), 145-168.
- Rodríguez, L.A. y Gamboa, M.E. (2019). Libro Excel para aplicar la escala de la resolución de problemas matemáticos (ResProMa). <http://roa.ult.edu.cu/jspui/handle/123456789/3960>
- Rodríguez, L.A., Gamboa, M.E. y Oliva, L.D. (2019). Diseño de escala de medición de la resolución de problemas del proceso de enseñanza-aprendizaje de la Matemática. En R. Velázquez (Presidencia), *Informática, Matemática y Ciencias de la Información*. Simposio llevado a cabo en 9na Edición de la Conferencia Científica Internacional de la Universidad de Holguín, Cuba.

- Rositas, J. (2014). Los tamaños de las muestras en encuestas de las ciencias sociales y su repercusión en la generación del conocimiento. *Innovaciones de negocios*, 11(22), 235-268.
- Solís, V. M. (2014). ¿Por qué algunos aún prohíben utilizar estadística paramétrica para analizar datos ordinales? *Enseñanza e Investigación en Psicología*, 19(2).
- Sánchez, R., y Echeverry, J. (2004). Validación de escalas de medición en salud. *Revista de Salud pública*, 6, 302-318.
- Stevens, S. (1946). On the Theory of Scales of Measurement. *Science, New Series*, 103(2684), 677-680.
- Velez, P. & Ashworth, S. D. (2007). The impact of item readability on the endorsement of the midpoint response in surveys. *Survey Research Methods*, 1(2), 69-74.
- Zapata, G. J. y Canet, M. T. (2008). Propuesta metodológica para la construcción de escalas de medición a partir de una aplicación empírica. *Actualidades investigativas en Educación*, 8(2), 1-26.
- Zeller, R.A. y Carmines, E.G. (1980). *Measurement in the social sciences. The link between theory and data*. Cambridge University Press.