

Introducción a la Estadística

Recopilar, describir, organizar e interpretar datos para transformarlos en información, nos permite tomar decisiones de una manera más eficiente. En este capítulo veremos cuáles son las formas de recopilar datos, distintos tipos de estudios estadísticos y métodos de análisis.

¿Qué es la Estadística?	14
Recopilación de datos	15
Estudios y métodos estadísticos	16
Algunas definiciones	18
Inicio de una investigación	21
Resumen	22
Actividades	24

¿QUÉ ES LA ESTADÍSTICA?

Con el objetivo de introducirnos en el estudio de la Estadística, le proponemos que lea las siguientes definiciones:

“El contenido de la estadística moderna incluye la recopilación, presentación y caracterización de la información a fin de que auxilie tanto en el análisis de datos como en el proceso de toma de decisiones”.⁽¹⁾

“La estadística, como campo de estudio, es el arte y la ciencia de dar sentido a los datos numéricos”.⁽²⁾

“La estadística se define a menudo como un conjunto de métodos para la toma de decisiones en condiciones de incertidumbre”.⁽³⁾

“La estadística estudia el comportamiento de los fenómenos llamados de colectivo. Está caracterizada por una información acerca de un colectivo o universo, lo que constituye su objeto material; un modo propio de razonamiento, el método estadístico, lo que constituye su objeto formal, y unas previsiones de cara al futuro, lo que implica un ambiente de incertidumbre, que constituyen su objeto o causa final”.⁽⁴⁾

“La estadística es el estudio de los fenómenos aleatorios. [...] El aspecto más importante de la estadística es la obtención de conclusiones basadas en los datos experimentales”.⁽⁵⁾

¿Qué palabras clave resaltaría de estas citas?

Aportamos algunas: **experimento, fenómeno, datos, información, métodos, técnicas, toma de decisiones, incertidumbre.**

(1) Berenson, M y Levine, D. *Estadística básica en administración*. Prentice-Hall Hispanoamericana, 4ta. Ed., 1992.

(2) Hildebrand, D. y Ott, L. *Estadística Aplicada a la Administración y a la Economía*. Addison-Wesley Iberoamerica, 3ra. Ed., 1997.

(3) Harnett, D. y Murphy, J. *Introducción al Análisis Estadístico*. Addison-Wesley Iberoamerica, 1987.

(4) Cabriá, 1994, en *Didáctica de la Estadística*, Batanero, P. 2001.

(5) Canavos, G. *Probabilidad y Estadística*. McGraw Hill, 1988.

Estas claves nos guiarán a través de las próximas páginas, en las que profundizaremos en temas tales como la obtención de datos y los métodos estadísticos que permiten la interpretación de los datos recopilados, haciendo hincapié en el hecho de que habitualmente nos enfrentamos a situaciones en las que debemos tomar decisiones, tanto en el ámbito laboral o profesional como en la vida cotidiana, y contamos sólo con información parcial, lo que nos coloca en un escenario de incertidumbre.

Recopilación de datos

Si bien la recopilación de datos es un tema que merece un párrafo aparte –sobre el que haremos algunas consideraciones en el **Capítulo 6**–, podemos anticipar que éstos pueden ser generados por el propio interesado, o bien haber sido recopilados por un tercero, ya sea un individuo o una organización. En el primer caso, la **fuerza de datos** se denomina primaria, y en el segundo caso, secundaria. En ambas situaciones, los datos pueden ser generados básicamente de tres maneras (**Figura 1**):

1. A través de la realización de un **censo**.
2. Tomando una **muestra**.
3. Diseñando un **experimento**.

Contar con datos censales implica que el interesado logró observar o medir todas las unidades o sujetos de interés existentes. Sin embargo, en muchas situaciones prácticas, no es posible acceder a todas las unidades existentes, por ser una tarea costosa, ardua y en ocasiones hasta imposible, o bien porque la medición destruiría a la unidad o sujeto en estudio.

¿Imagina por ejemplo un censo sobre los insectos de cierta especie? En tal caso será más conveniente –cuando no, el único camino– obtener los datos a través de la medición u observación de una porción de las unidades existentes, es decir, tomando una muestra.

El tamaño de la población a estudiar juega un papel crucial al momento de decidir de qué manera se han de recopilar datos: si la población es infinita o finita pero de gran tamaño, la muestra será la manera más razonable y eficiente de recopilar datos. O, ¿cómo controlaría la calidad de los envases de vidrio salidos de la producción de una pequeña fábrica? Si se trata del aspecto estético, podría contar las burbujas de cada envase para clasificar cada pieza pro-

ducida. Si, en cambio, deseara estudiar la resistencia de la materia prima, ¿intentaría usted romper cada pieza fabricada? Ya sea por ahorro de tiempo o bien por pruebas destructivas una muestra de piezas para control de calidad es la forma de recopilación más razonable en muchos casos.

En otros casos es necesario que el investigador genere sus propios datos interviniendo directamente, por lo que será preciso que diseñe un experimento a la medida de sus necesidades.

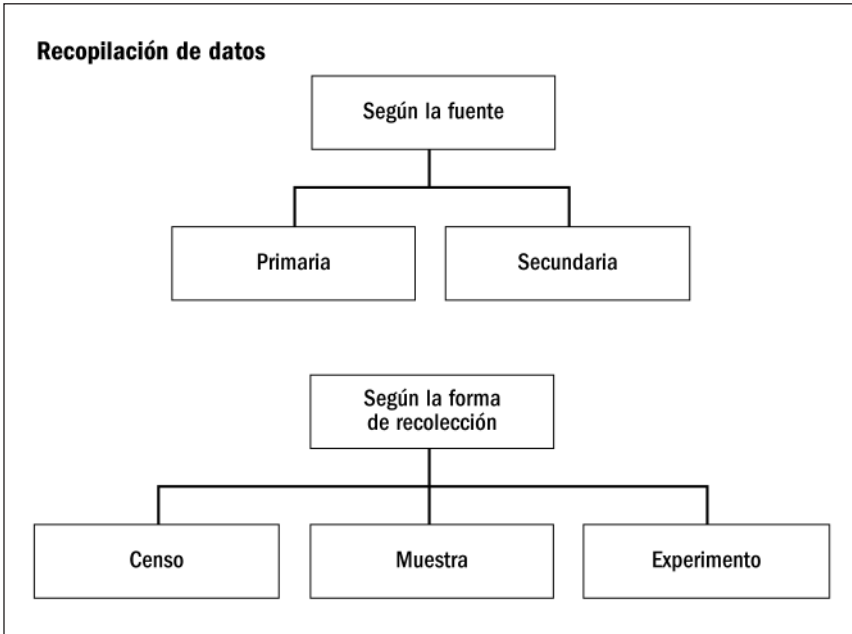


Figura 1. ¿Cómo reunir datos?

Antes de seguir adelante, le sugerimos que realice la **Actividad 1** que encontrará al final del capítulo, con la que podrá aplicar los conceptos hasta aquí expuestos.

Estudios y métodos estadísticos

Ya sea que los datos se hayan generado a partir de un censo o una muestra, podemos realizar **estudios observacionales**, midiendo u observando las unidades o sujetos y registrando las características de interés.

Los estudios observacionales pueden ser retrospectivos o prospectivos, según evalúen situaciones pasadas en busca de una explicación en acontecimientos actua-

les, o bien realicen un seguimiento en el futuro de los efectos de alguna circunstancia observada en el presente o en el pasado. Un ejemplo de estudio retrospectivo es la recopilación de información sobre la capacitación y estilo de liderazgo de ejecutivos que, actualmente, se reconocen como exitosos. Un ejemplo de estudio prospectivo es el registro sistemático de información sobre la capacitación y estilo de liderazgo de actuales profesionales novatos con el fin de establecer el grado de éxito profesional que alcanzan en el momento de su retiro.

Si, en cambio, el investigador diseña un experimento a medida, que consiste en aplicar uno o más tratamientos a las unidades o sujetos, y observa la respuesta, se encaminará hacia un tipo de **estudio experimental**.

Por otra parte, el objetivo perseguido por el analista puede dar lugar a otra tipología de estudios estadísticos. Ellos son los **estudios exploratorios** y los **confirmatorios** (Figura 2). Los primeros se aplican cuando se desea buscar un modelo que explique o describa convenientemente el comportamiento de alguna variable, la relación entre variables o grupos de sujetos, etc. En cambio, los estudios confirmatorios se orientan a la verificación de la compatibilidad de un conjunto de datos con respecto a un modelo supuesto previamente.



Figura 2. ¿Qué variedad de estudios estadísticos hay?

Para llevar adelante alguno de los estudios mencionados, la Estadística nos provee de una variedad de métodos que podríamos clasificar en **univariados** y **multivariados**, dependiendo de la cantidad de variables que se analizan simultáneamente (**Figura 3**).

Si nos ocupa el análisis de uno o varios conjuntos de datos de la misma variable, contamos con métodos gráficos y numéricos que reúne la rama de la Estadística que se conoce como **estadística descriptiva** (que trataremos en el **Capítulo 2**), o bien aquella rama conocida como **estadística inferencial**, que facilita la toma de decisiones bajo condiciones de incertidumbre a través de las pruebas de hipótesis y las estimaciones (esta última será desarrollada en el **Capítulo 6**).

Si nos ocupa el análisis de dos conjuntos de variables denominadas explicativas y explicadas, nos serán de utilidad los métodos de regresión (alguno de los cuales se podrá consultar en el **Capítulo 3**); mientras que los métodos factoriales nos permiten analizar un único conjunto de variables explicativas como el análisis de clusters, de componentes principales y análisis discriminante, entre otros (técnicas que están fuera del alcance de este texto).

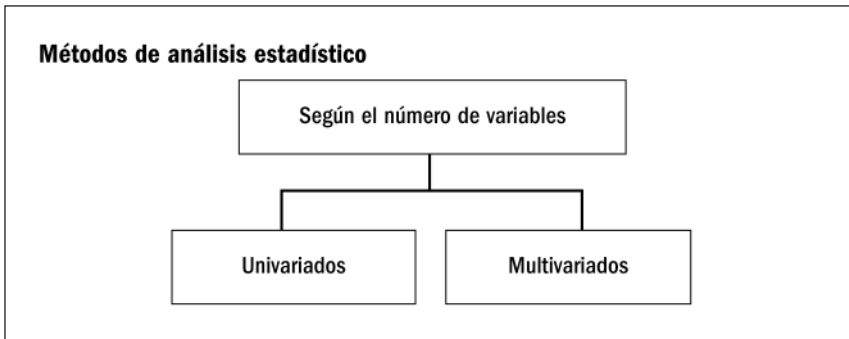


Figura 3. ¿Qué métodos de análisis de datos ofrece la Estadística?

La **Actividad 2** propuesta al final del capítulo está diseñada para que aplique los temas tratados en este parágrafo.

ALGUNAS DEFINICIONES

En las páginas previas hemos utilizado algunos vocablos que, si bien pueden resultar familiares, es conveniente definir con una mayor precisión.

Con este objetivo, repasemos los procedimientos básicos de recopilación de datos. Mencionamos oportunamente que tales procedimientos eran tres: los censos, las muestras y los experimentos. Una pregunta básica que hemos de hacernos antes de decidir entre ellos es ¿cuál es la población de interés? Una de las acepciones de la palabra **población** se refiere al conjunto de unidades elementales con características comunes observables.

Si, por ejemplo, nos proponemos conocer la frecuencia con que los clientes de un salón de belleza demandan algún servicio, razonamos que la población en estudio se compone de “todos los clientes actuales” del negocio, que quizá representen para nosotros el conjunto de sujetos que se atendieron en el salón de belleza al menos dos veces en el último año. Por supuesto, una **muestra** es un subconjunto de unidades elementales tomadas de la población.

Por su parte, una **unidad elemental** es cada objeto o sujeto que observamos de la población. En nuestro ejemplo, una unidad elemental es cada individuo que se atendió en el salón de belleza dos veces o más en los últimos doce meses.

Si a cada persona que requirió los servicios del salón de belleza se le completa una ficha donde consta la fecha de cada visita, podremos fácilmente identificar a aquellos que forman parte de la población. Concretamente observaremos la cantidad de veces que cada cliente tomó algún servicio en el año: si es 1 o menos, lo descartaremos; si es 2 o más, será considerado con atención a los fines del estudio, porque forma parte de la población bajo análisis.

Aquella característica que se observa o se mide sobre las unidades elementales se denomina **variable**.

Por supuesto, podríamos reunir datos sobre otras variables relacionadas con los clientes del salón de belleza. Contando con un fichero actualizado, podríamos recopilar información relevante como, por ejemplo, qué servicios tomó cada cliente en cada visita o el gasto que efectuó.

Observe que las variables del ejemplo son bastante diferentes entre sí; en realidad son de distinto tipo y se miden en diferente escala.

El “tipo de servicio tomado por cada cliente” se clasifica como **variable cualitativa o categórica**; no toma valores numéricos, sino que asume categorías como podrían ser “tratamiento capilar”, “tratamiento dermatológico”, “manicura”, etc. La variable de este ejemplo en particular está medida en **escala nominal**, por cuanto cada unidad de la escala representa una categoría diferente y entre ellas no guardan una relación de magnitud. En otras palabras, pa-

ra la variable “tipo de servicio tomado por cada cliente”, la categoría “tratamiento capilar” no es más ni menos importante que la categoría “tratamiento dermatológico”. Otros ejemplos de variables medidas en escala nominal son las distintas marcas de gaseosa o las nacionalidades.

Si, en cambio, las categorías de una variable guardan un cierto orden de importancia, como “tratamiento sencillo”, “tratamiento vip”, tal variable está medida en **escala ordinal**. En esta escala de medición es posible establecer relaciones del tipo $A > B$, $A = B$ o $A < B$, aunque no informa acerca de la magnitud de las diferencias entre categorías. No podríamos afirmar que el tratamiento vip es el doble de importante que el sencillo, o bien que un alumno excelente es el triple de bueno que un alumno regular.

Debemos saber que aquellas variables que se midan en escala nominal u ordinal no nos permitirán realizar cálculos matemáticos y que, a lo sumo, contaremos con el número de individuos que se clasifican en una u otra categoría. Por otra parte, si consideramos la variable “cantidad de veces que cada cliente se atendió en el salón de belleza en el último año”, observamos que puede asumir valores numéricos, aunque no cualquier valor. Se trata de una **variable cuantitativa**, que a su vez se subclasifica como **discreta** para el ejemplo. Se define como aquella variable que no puede ser particionada infinitamente dentro de un intervalo. Si es nuestro interés analizar “el monto pagado en cada visita al salón de belleza”, también clasificaremos la variable como cuantitativa; sin embargo, razonamos que dentro de un rango, en realidad hay infinitos valores posibles, por lo tanto se trata de una **variable cuantitativa continua**. Este tipo de variables puede ser particionado infinitamente dentro de un intervalo.

Las variables cuantitativas pueden estar medidas en **escala de intervalos** o en **escala de razón**. La escala de intervalos representa un nivel superior de medición con respecto al de la escala ordinal. Además de permitir relaciones del tipo $A > B$, $A = B$, o $A < B$, es posible comparar intervalos de valores debido a que a lo largo de toda la escala, dos valores adyacentes representan siempre la misma diferencia de magnitud, por lo que es posible realizar comparaciones del tipo $A - B = C - D$, $A - B < C - D$, o $A - B > C - D$. Una particularidad de esta escala es que el valor cero es arbitrario. Un ejemplo de aplicación de esta escala lo representan los horarios de llegada de los empleados de una compañía. Entre las 8 y las 9 am existe la misma diferencia que entre las 4 y las 5 pm, pero no puede afirmarse que quien llegó a las 8 pm haya arribado el doble de tarde que el que ingresó a las 8 am. Incluso la hora cero no indica especialmente nada, como “ausencia del empleado”.

El nivel más alto de medición es la escala de razón. Además de tener las propiedades de las otras escalas, tiene un cero absoluto, por lo tanto es posible calcular proporciones entre los valores de la escala. Así puede establecerse que, por ejemplo, un individuo de 40 años tiene el doble de edad de uno de 20 años, que entre 20 y 21 años existe la misma diferencia etaria que entre 40 y 41 años, y también que un sujeto de 40 años es mayor que uno de 20. De manera que sobre esta escala es posible realizar todas las operaciones matemáticas asociadas a los números (suma, resta, multiplicación y división). Relaciones similares podrían realizarse sobre variables como la remuneración, la estatura o el precio de cierto artículo, que siendo variables cuantitativas, en ocasiones y por cuestiones prácticas, pueden “discretizarse” y hasta “categorizarse” modificando su escala de medición original. La correcta clasificación de las variables en estudio, así como la definición de la escala de medición, orientará al investigador sobre las técnicas estadísticas más adecuadas para analizarlas.

En páginas anteriores anticipamos que la palabra **población** tiene dos acepciones. Mencionamos una de ellas para la que “población” representa el conjunto completo de objetos o sujetos bajo análisis. La otra acepción del vocablo se relaciona con el conjunto completo de valores que toma una variable en un universo. Tal conjunto puede ser descrito y resumido con una serie de medidas descriptivas denominadas **parámetros**, mientras que a las medidas descriptivas de datos muestrales se las llama **estimadores** o **estadígrafos**.

Le proponemos realizar la **Actividad 3**, con la que usted podrá poner en práctica los conceptos del presente capítulo.

Inicio de una investigación

El éxito de una investigación radica en buena parte en el cuidadoso diseño previo al trabajo de campo orientado a la obtención de datos. Para ello el investigador debe tener respuestas claras a cuestiones trascendentes como las que siguen:

¿Cuál o cuáles son los objetivos de la investigación?

- ¿Resolver un problema? ¿Se tienen soluciones provisionarias?
- ¿Tomar decisiones? ¿Cuáles son los caminos de acción posibles?
- ¿Incrementar el conocimiento sobre algún tema de interés?
- ¿Adhiere el investigador a alguna teoría conocida?
- ¿Hay un objetivo general y uno o más objetivos específicos?

¿Qué tipo de estudio se llevará a cabo?

- ¿Cuál es la población de interés?
- ¿O bien se trata de comparar dos o más poblaciones?
- ¿Cuáles son las unidades elementales?
- ¿La población es infinita o finita? ¿Es accesible?

¿Qué variables pueden proveer la información para satisfacer el objetivo?

- ¿Qué instrumento de medición es más adecuado para medir las variables?
- ¿En qué escala serán medidas?

¿Cómo se recopilará la información?

- ¿Hay información disponible, accesible y actualizada, o bien se la debe recoger o generar?
- ¿La accesibilidad y los recursos disponibles son suficientes para hacer un censo?
- ¿Qué método de selección de muestras es más apropiado?
- ¿De qué tamaño será la muestra?
- ¿Qué diseño de experimento se adecua más a los recursos disponibles?

¿Qué herramientas estadísticas son aplicables al caso?

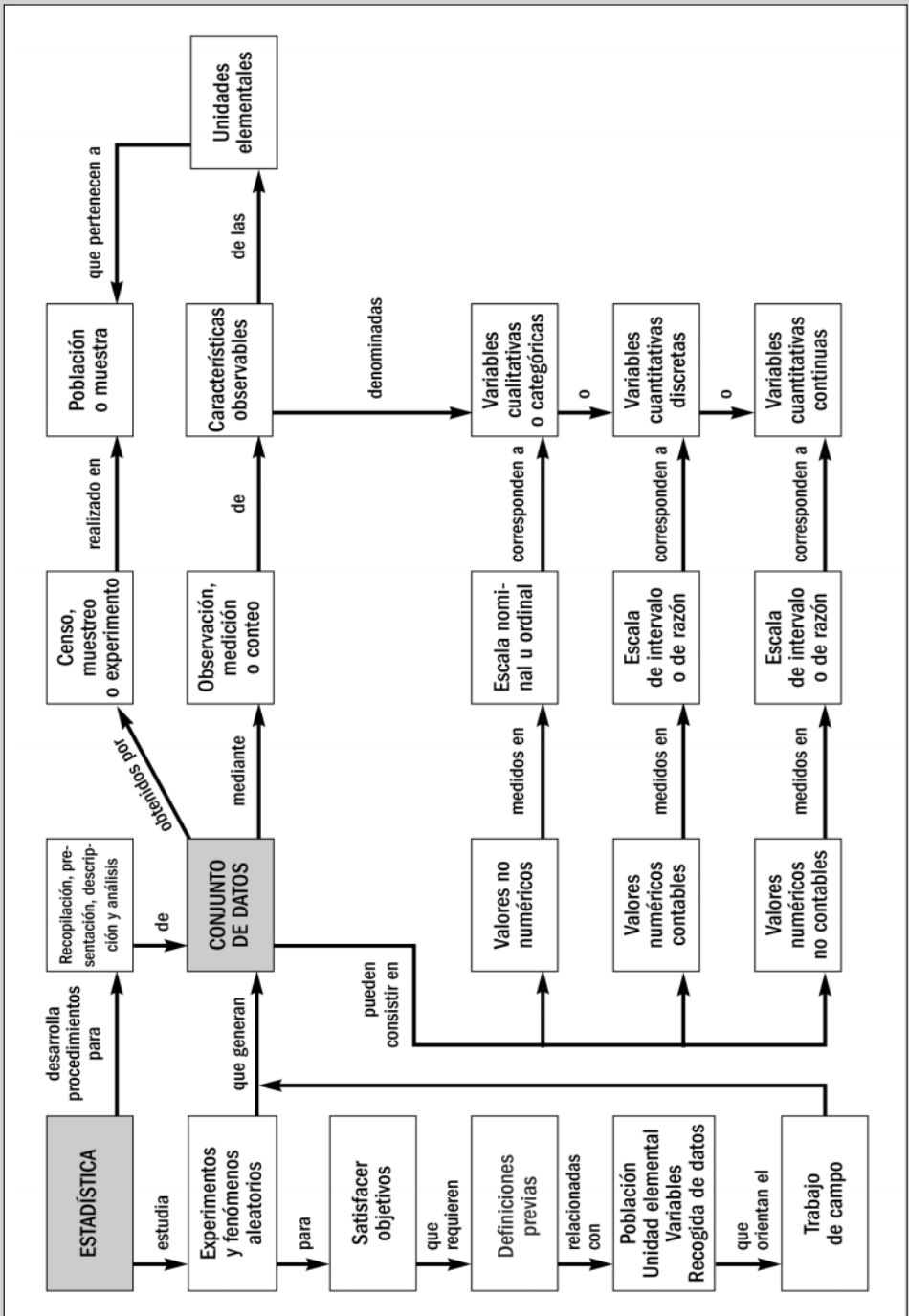
- ¿Se estudia una o más variables? ¿Cómo se clasifican?
- ¿Se estudia una o más poblaciones?

Habrás notado el lector que al iniciar una investigación en cualquier campo –científico, comercial, académico, etc.– resulta imprescindible contar con ciertos conceptos elementales: como mínimo, los que presentamos en este capítulo.

Podríamos continuar la lista, pero apostamos a que, a medida que el lector avance en la lectura, sea capaz de enriquecerla por sí mismo.

RESUMEN

Seguramente usted tenga la habilidad de integrar los conceptos que abordamos en este capítulo. Por nuestra parte le facilitamos un diagrama, que se encuentra en la página siguiente, que tiene el objetivo de resumir los puntos tratados de manera esquemática.





ACTIVIDADES

- 1** Para las siguientes propuestas, decida de qué manera recopilaría los datos y a qué tipo de fuente recurriría:
 - a. Objetivo: estudiar la evolución de la cotización de las acciones de cierta fábrica automotriz en el último año.
 - b. Objetivo: conocer la frecuencia de compra de los clientes de su propio negocio.
 - c. Objetivo: analizar los efectos de un nuevo analgésico.

- 2** Para las siguientes propuestas, decida qué tipo de estudio estadístico efectuaría y qué método/s podría aplicar para el análisis de datos:
 - a. Usted es un comerciante y desea estudiar el comportamiento de las ventas en su negocio en función de su inversión en publicidad y la cotización del dólar.
 - b. En una investigación sobre tabaquismo se desea analizar los efectos a largo plazo de tal hábito sobre el sistema respiratorio.
 - c. Un fabricante de ropa deportiva desea conocer el perfil de sus clientes.

- 3** Lea las siguientes variables e intente identificar la población en estudio, las unidades elementales, su unidad de medida y su escala de medición:
 - a. Lugar de origen de los pasajeros de cierta empresa de aviación arribados en el último vuelo al aeropuerto internacional de Barajas.
 - b. Número de canciones de un CD.
 - c. Temperatura máxima diaria en Caracas durante el último mes.
