

Investigación Estadística



Ing. Jorge Aguirre Orozco, MBA.

Estadística en las ciencias

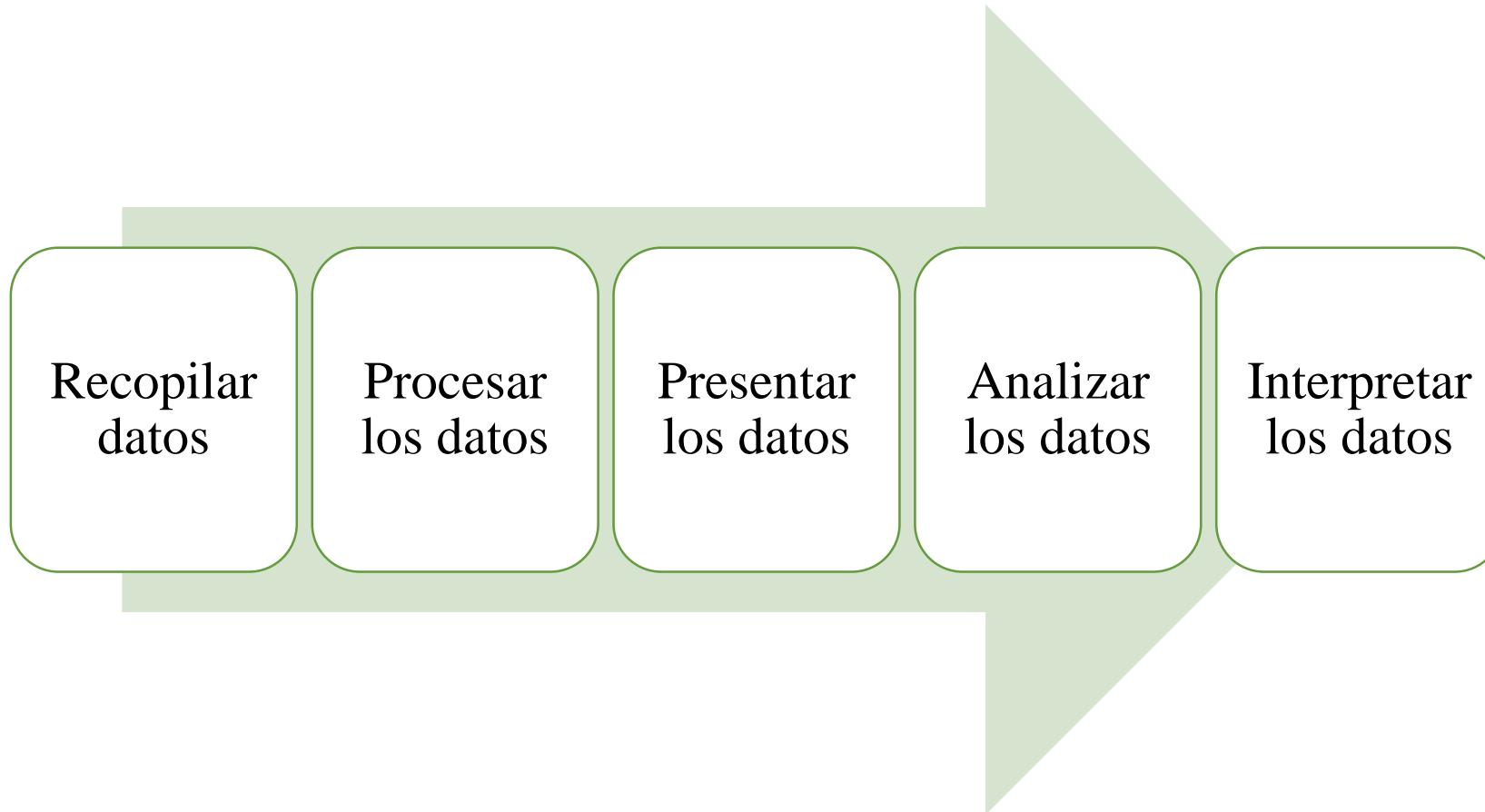
Las matemáticas han adquirido gran importancia en las ciencias exactas y en las ciencias sociales, ya que estas últimas recurren a la estadística como una herramienta para medir y cuantificar fenómenos sociales.

Interpretación de la estadística

“El método que permite organizar, sintetizar, presentar, analizar, cuantificar e interpretar una gran cantidad de datos, de tal forma que se puedan tomar decisiones, realizar generalizaciones y obtener conclusiones válidas (dar información) sobre los fenómenos o líneas de investigación en estudio.”



Funciones de la estadística



Ramas de la estadística

Estadística Descriptiva:

“Se encarga de obtener, organizar, presentar y describir **los datos mediante la aplicación de métodos y técnicas.** Aplica exclusivamente a los datos que forman una muestra, lo cual permitiría inferir o generalizar acerca de la totalidad de los elementos que constituyen la población de estudio.”

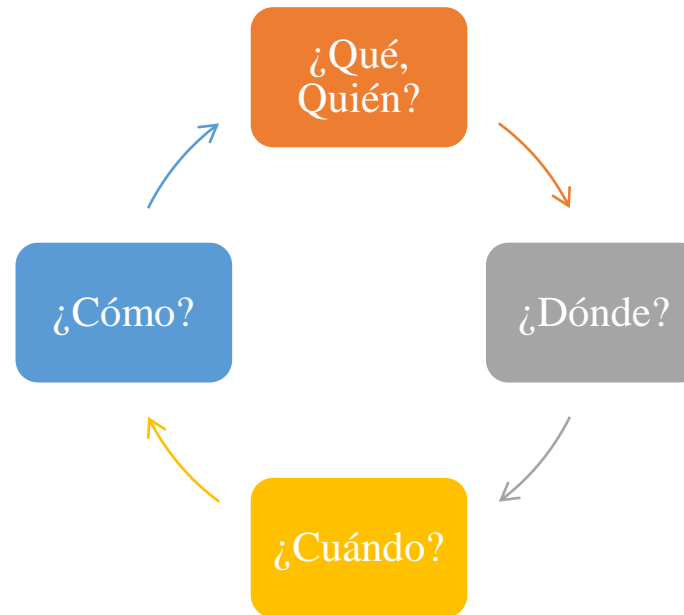
Ramas de la estadística

Estadística Inferencial:

“Sus objetivos son generalizar o deducir a partir de estudios de muestra, el comportamiento de una población; con ello se pueden tomar decisiones útiles. La información que proporcionan las muestras se obtiene de la estadística descriptiva.”

Unidad estadística

Es aquella unidad u objeto de interés en la investigación y de ella se desprenden las observaciones, o sea, de ella se derivan los datos para el análisis.



Población

Es el conjunto formado por un número determinado o indeterminado de unidades (personas, objetos, fenómenos, etc.) que comparten características comunes a un objeto de estudio.

- *Población finita*: elementos finitos que pueden contarse.
- *Población infinita*: elementos que no pueden listarse o enumerarse en su totalidad.

Muestra

Es cualquier subconjunto seleccionado de una población, que sigue ciertos criterios establecidos en la teoría del muestreo. La muestra es el elemento básico en el cual se fundamenta la posterior inferencia acerca de la población de donde procede.



Muestra

- *Aleatoria*: se emplea un procedimiento que conceda unidad una probabilidad conocida de formar parte de la muestra. Ejemplo: azar, sistemático, estratégico, de conglomerados (implican sesgos e imposibilidad de hacer inferencias).
- *No Aleatoria*: se emplea un procedimiento que no es posible calcular la probabilidad de que cada unidad sea seleccionada en la muestra. Ejemplo: por conveniencia, de juicio, voluntario, otros.

Muestreo aleatorio

MUESTREO ALEATORIO SIMPLE

Técnica de muestreo en la que cada miembro de la población tiene la misma probabilidad de ser seleccionado en la muestra.

=ALEATORIO.ENTRE (1;10)							
B	C	D	E	F	G	H	I
ID	NOMBRE	CEDULA	PROVINCIA				
1	JOSE CASTRO PEREZ	3-508-687	CARTAGO				
2	TERESA MENDEZ CORDOBA	7-006-956	LIMON				
3	ROCIO SANCHEZ PEREZ	1-1033-0456	SAN JOSE				
4	CLAUDIO SANDI LOPEZ	5-087-054	GUANACASTE				
5	KARLA MONGE TRIGUERO	4-056-799	HEREDIA				
6	ESTER ESPINOZA PAZ	2-987-356	ALAJUELA				
7	LUCIA PORRAS GUZMAN	1-1133-0555	SAN JOSE				
8	SEBASTIAN PEREZ COT	3-734-890	CARTAGO				
9	HUGO AVALOS PORRAS	4-345-521	HEREDIA				
10	JUAN AGUILAR LONDOÑO	7-345-000	LIMON				

n=1	1
n=2	6
n=3	2
n=4	7

Muestreo aleatorio

**MUESTREO ALEATORIO
ESTRATIFICADO** Técnica de
muestreo que divide la población en
subgrupos, denominados estratos,
y selecciona al azar una muestra de
cada estrato.

Estrato	Antigüedad	Número de empleados	Frecuencia relativa	Tamaño de Muestra
1	menos de 5 años	200	0,20	10
2	De 5 a 10 años	500	0,50	25
3	10 años o más	300	0,30	15
	Total	1000	1,00	n= 50
	Estrato 1:	$=(0,2*50)$	10	
	Estrato 2:		25	
	Estrato 3:		15	

Muestreo aleatorio

MUESTREO ALEATORIO SISTEMÁTICO

Técnica de muestreo que selecciona un punto aleatorio de inicio y posteriormente se elige cada k -ésimo elemento de la población.

Constante de espaciamiento

$$k = \frac{N}{n}$$

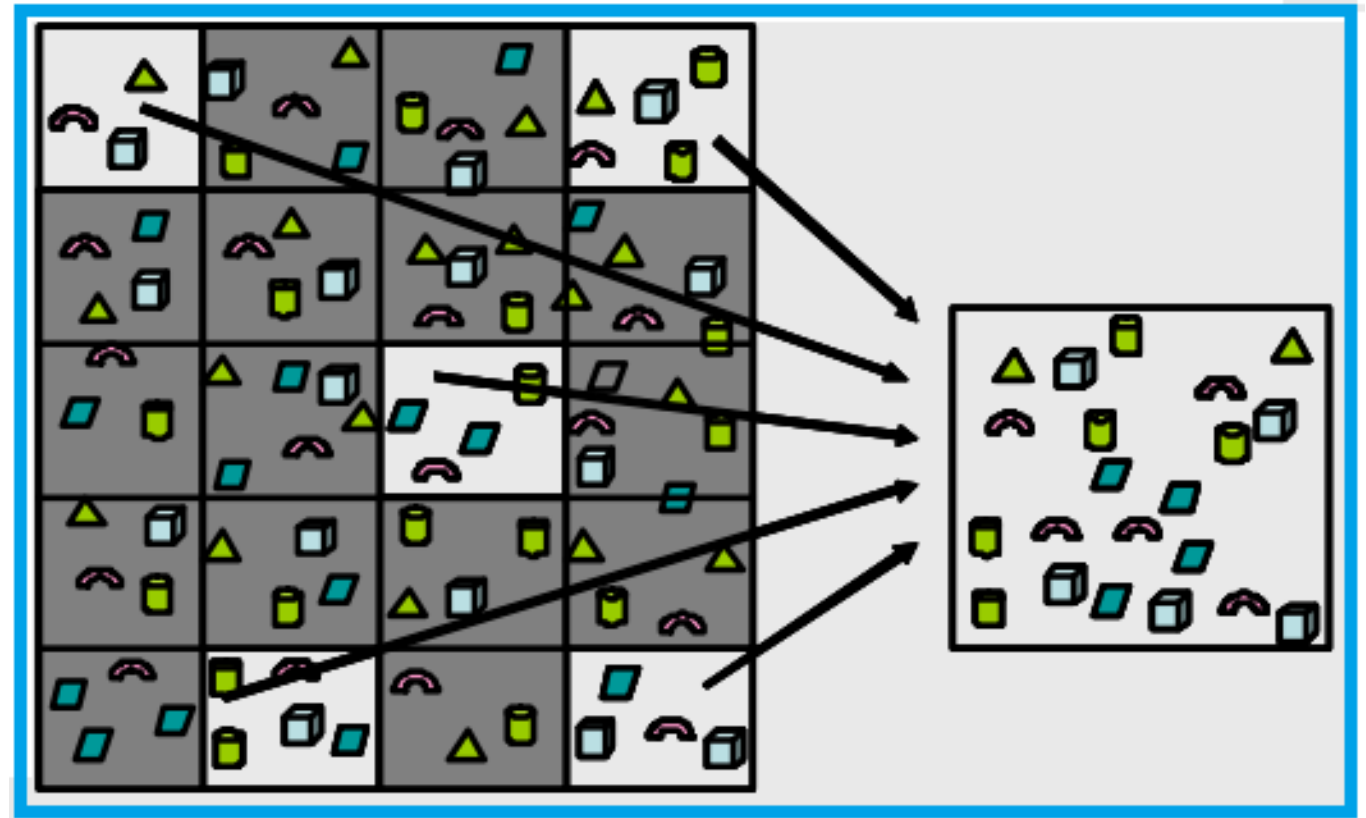
N=	10
n=	5
k=	2
Elemento de la muestra	Elemento seleccionado de la población
1	2
2	2 + 2 = 4
3	4 + 2 = 6
4	6 + 2 = 8
5	8 + 2 = 10

Valor aleatorio

Muestreo aleatorio

MUESTREO ALEATORIO POR CONGLOMERADOS

Técnica de muestreo que divide la población en conglomerados a partir de sus límites naturales geográficos o de otra clase, para luego seleccionar aleatoriamente conglomerados y tomar una muestra al azar con elementos de cada uno de ellos.



Muestreo no aleatorio

MUESTREO DE JUICIO en este tipo de muestreo las unidades son seleccionadas mediante el juicio personal del investigador. Usualmente es experto en la materia.

OTROS TIPOS DE MUESTREO NO ALEATORIO en ocasiones se habla de del muestreo por cuotas, de las muestras autogeneradas, entre otras formas. Este tipo no permite establecer la representatividad de la muestra de forma objetiva.

MUESTREO POR CONVENIENCIA en este caso el investigador selecciona para conformar la muestra, aquellas unidades que sean más fáciles de acceder. En otras palabras, se elige de acuerdo con la comodidad del investigador.

Errores de muestreo y sesgos

- *Error de muestreo:*

Es la diferencia causada por el azar, entre el valor verdadero de una característica y el valor estimado mediando el uso de una muestra.

- *Sesgo:*

Error sistemático, no debido al azar, entre el valor verdadero de una característica y el valor estimado en una investigación.

Variable

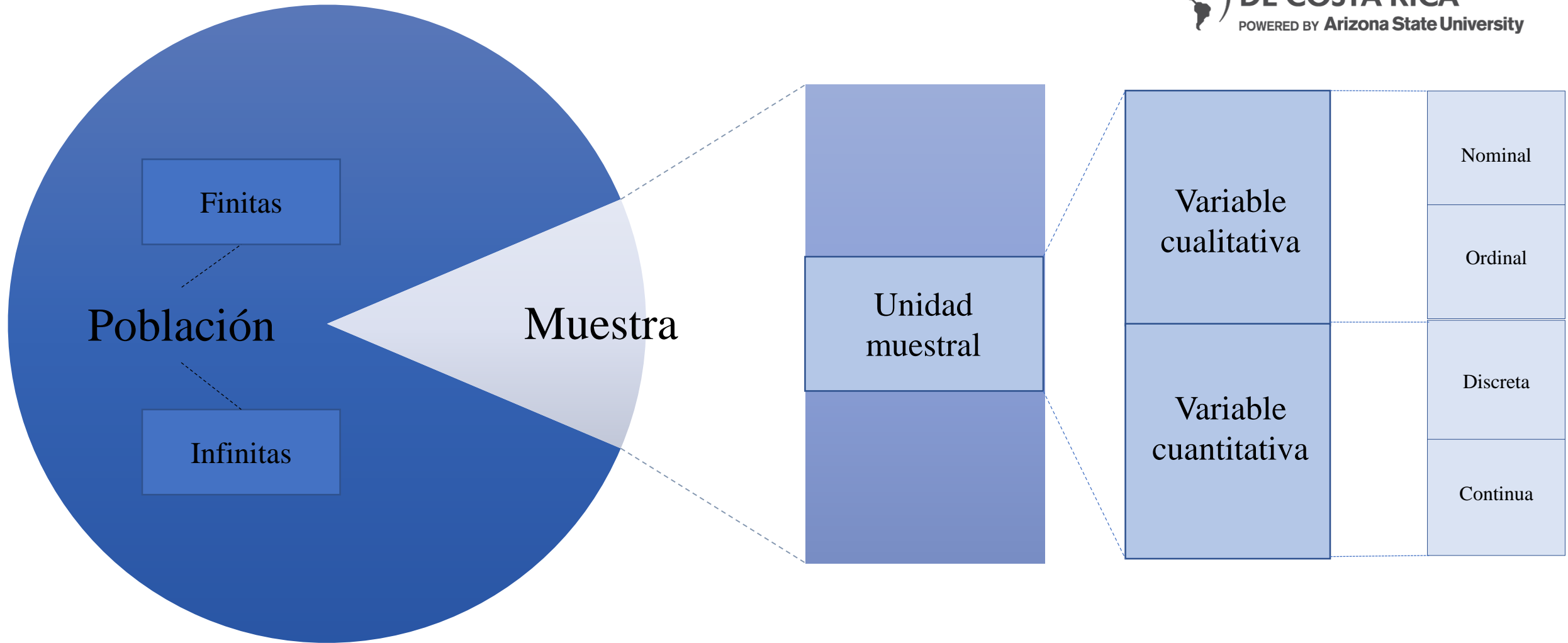
- Para obtener estadística se utilizan conjuntos que poseen un determinado o indeterminado número de unidades, como son: los objetos, las personas, los fenómenos, etc.
- Cada una de las unidades de estudio tienen determinadas características; ejemplo, para el salón de la clase se puede estudiar: la edad, estado civil, estatura, género, grado de escolaridad, lugar de nacimiento, peso, la talla, entre otras.

Tipos de variable

- Variable cualitativas:
 - Nominal: modalidades no numéricas que no admiten un criterio de orden. Ejemplo: estado civil con las siguientes modalidades: soltero, casado, separado, divorciado y viudo.
 - Ordinal: modalidades no numéricas, en las que existe un orden. Ejemplo: medallas de una prueba deportiva: oro, plata, bronce.

Tipos de variable

- Variables cuantitativas:
 - Discreta: solo puede tomar un número finito de valores entre dos valores cualesquiera de una característica. Ejemplo: número de hermanos de 5 amigos: 2,1,0,1,3.
 - Continua: puede tomar un número infinito de valores entre dos valores cualesquiera de una característica. Ejemplo: la altura de los 5 amigos: 1.73,1.82,1.77,.1,69,1.75 (*instrumento de medición*).



Técnicas de recolección de datos

Censo

- Estudio de todos y cada uno de los elementos de una población.

Éste consiste en aplicar un cuestionario a todos los elementos que integran una población; es un estudio exhaustivo de la población.

Nota: es opción cuando las poblaciones son pequeñas, se requiere una exactitud completa o el tamaño de la muestra se acerca al tamaño de la población. No es opción cuando, se requiere tener información rápidamente o hay escasez de recursos.



Métodos e instrumentos de recolección de información

OBSERVACIÓN consiste en que el investigador recopila y registra datos observando (midiendo o contando) aquellas características que son de interés para los objetivos de la investigación (se apoya en sus sentidos (vista, oído, olfato, tacto..))





Métodos e instrumentos de recolección de información

ENTREVISTA PERSONAL consiste en la visita del investigador a las unidades de estudio y les plantea una serie de preguntas, las cuales se registran en un cuestionario previamente diseñado para este fin.

ENTREVISTA TELEFÓNICA es similar al método anterior, la diferencia consiste en que los informantes son contactados por medio de una llamada telefónica.



Métodos e instrumentos de recolección de información

CORREO en este caso se emplea un cuestionario, pero se envía por correo postal o electrónico y se recibe por el mismo medio. Actualmente existe software que facilita la aplicación de encuestas a través de internet.

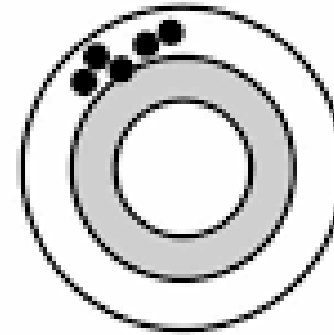
REGISTRO si el investigador conoce el dato exacto sobre una variable, no es necesario aplicar alguno de los métodos de recolección de información; en ocasiones existen organismos que registran los mismos.



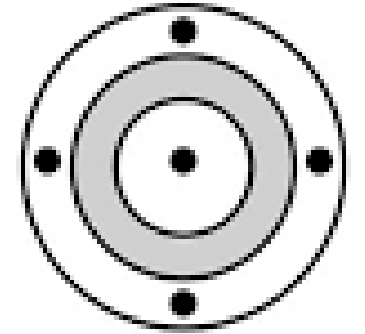
Métodos e instrumentos de recolección de información

Confiabilidad: se relaciona con la capacidad del instrumento de dar medidas “certeras”, es decir, que se acerquen a la verdadera medida de la variable considerada.

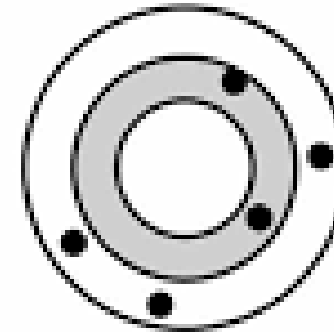
Validez: se refiere a que el instrumento mida realmente lo que se quiere medir. Para que un instrumento sea válido debe ser confiable, sin embargo, podría ser confiable sin ser válido.



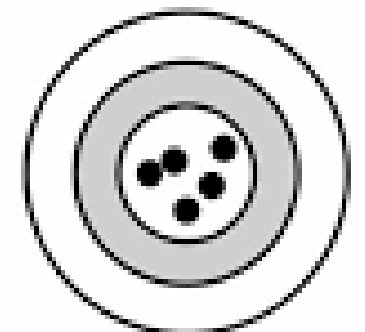
Confiable, poca validez



Válido, pero poco confiable



Sin validez ni confiabilidad



Válido y confiable

Elementos de la investigación

- *Formulación del problema:* consiste en identificar y especificar adecuadamente un problema de investigación. En esta etapa es muy importante establecer con precisión la o las hipótesis, el o los objetivos del estudio, su alcance y la población de datos asociada con el mismo.
- *Diseño de experimento:* consiste en seleccionar la técnica de recolección de datos (observación directa, entrevista, encuesta) que le permita obtener información a un mínimo costo (dinero y tiempo) posible. También definir el tamaño de la muestra, calidad requerida y el tipo de datos que le permitan resolver el problema planteado.

Elementos de la investigación

- *Recolección de datos*: etapa de mayor importancia en la investigación, ya que la calidad de los datos obtenidos depende de una óptima recolección; la cual debe sujetarse a reglas estrictas que permitan obtener la información deseada. Con frecuencia es necesario emplear un instrumento de recolección de datos que sea válido y confiable.
- *Proceso de datos y su descripción*: consiste en elaborar cuadros estadísticos de referencia, gráficas y cálculos de medidas estadísticas apropiadas al proceso descriptivo o inferencial seleccionado. Exponer los datos mediante diferentes instrumentos con el objeto de hacer una descripción de los resultados.

Elementos de la investigación

- *Inferencia estadística y conclusiones*: proporciona una contribución muy importante, ya que en ella se define el nivel de confianza y significación del proceso inferencial, lo cual sirve como orientación a quien o quienes deben tomar una decisión acerca del tema objeto de estudio. Conlleva el tratamiento matemático de la información que permita realizar comparaciones y establecer relaciones.
- *Presentación de informe*: se busca comunicar aspectos como:
 - Problema que dio origen a la investigación.
 - Hipótesis planteada y cómo se puso a prueba.
 - Metodología aplicada en la recolección de información y análisis de los datos.
 - Resultados obtenidos.
 - Conclusiones finales.

Presentación de los datos

Formas de presentación de la información estadística



TEXTUAL:

Los datos van descritos dentro del texto:

“El 10% de los entrevistados está de acuerdo con la política del gobierno, mientras que el 90% manifiesta estar en desacuerdo”



TABULAR:

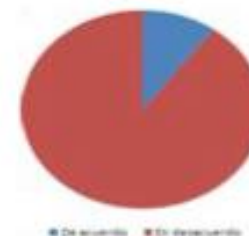
Los datos se presentan usando tablas o cuadros:

Opinión	Porcentaje
De acuerdo	10%
En desacuerdo	90%



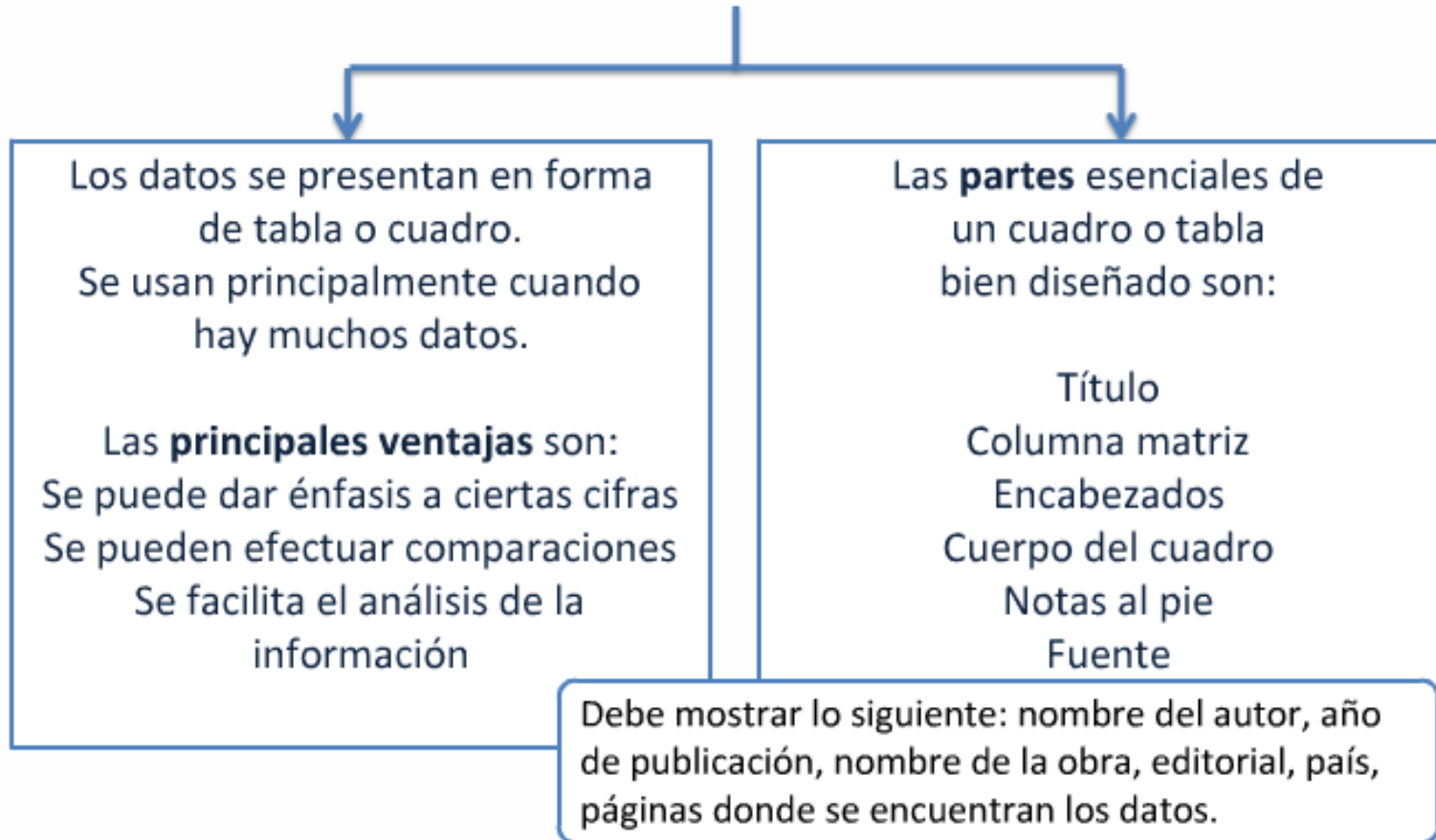
GRÁFICA:

Los datos se presentan usando gráficas:



Presentación de los datos

Presentación tabular



Presentación de los datos

Ejemplo Presentación tabular

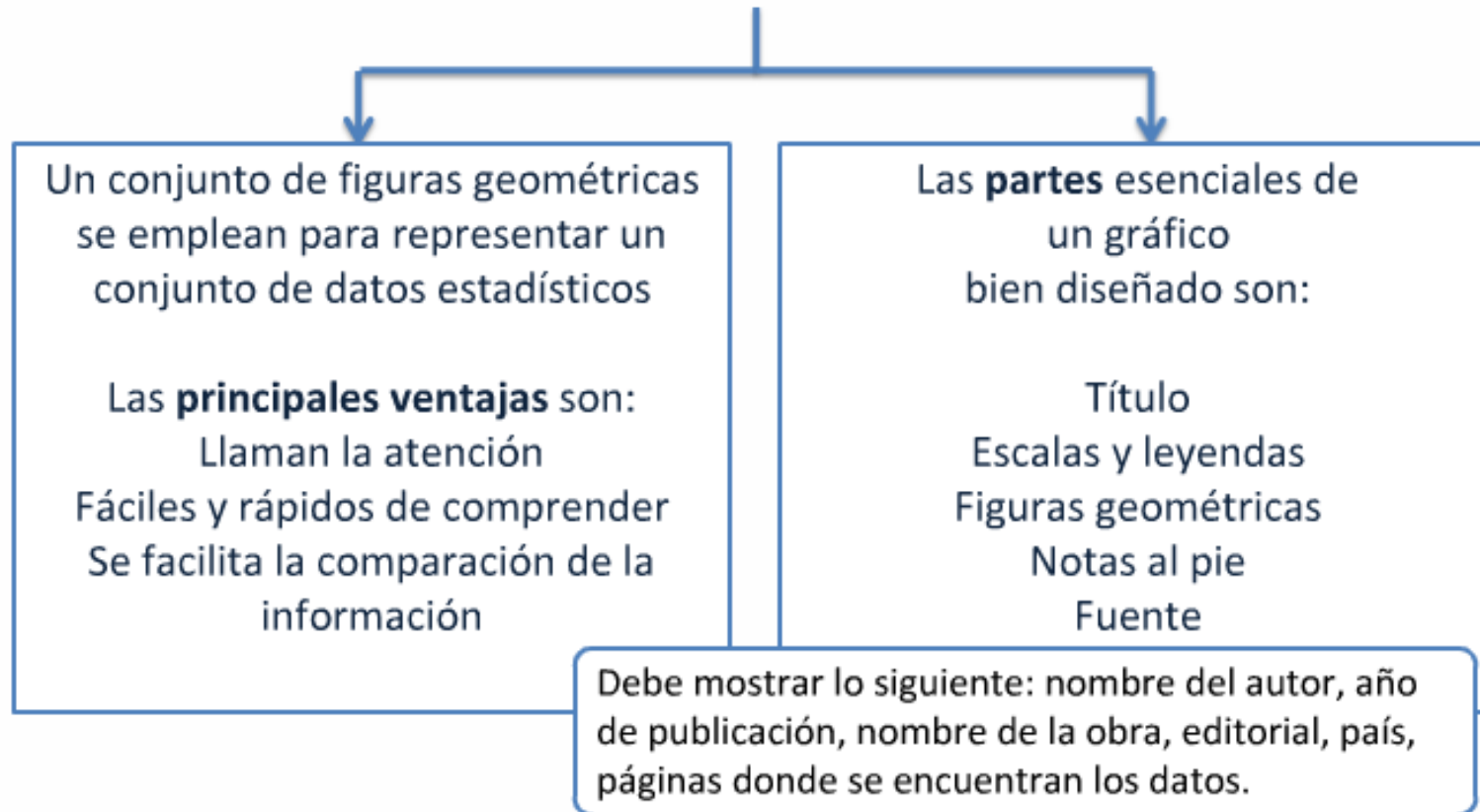
Cuadro #1
**Centroamérica: matrícula inicial en los primeros niveles,
según nivel educativo, por países.
1997
(datos en miles)**

País	Nivel			Total
	Preescolar	Primaria	Secundaria	
Costa Rica	71.1	525.3	220.2	816.5
El Salvador	179.0	1191.1	152.5	1522.6
Guatemala	221.8	1539.5	317.4	2078.6
Honduras	86.1	1055.0	271.1	1412.1
Nicaragua	133.1	777.9	270.5	1811.5
Panamá	45.3	377.9	223.2	646.4
Total	736.4	5466.6	1451.7	7657.7

Fuente: Informe Estado de la Región en Desarrollo Humano Sostenible.
Proyecto Estado de la Región. San José, C.R.: 1999, pág. 169.

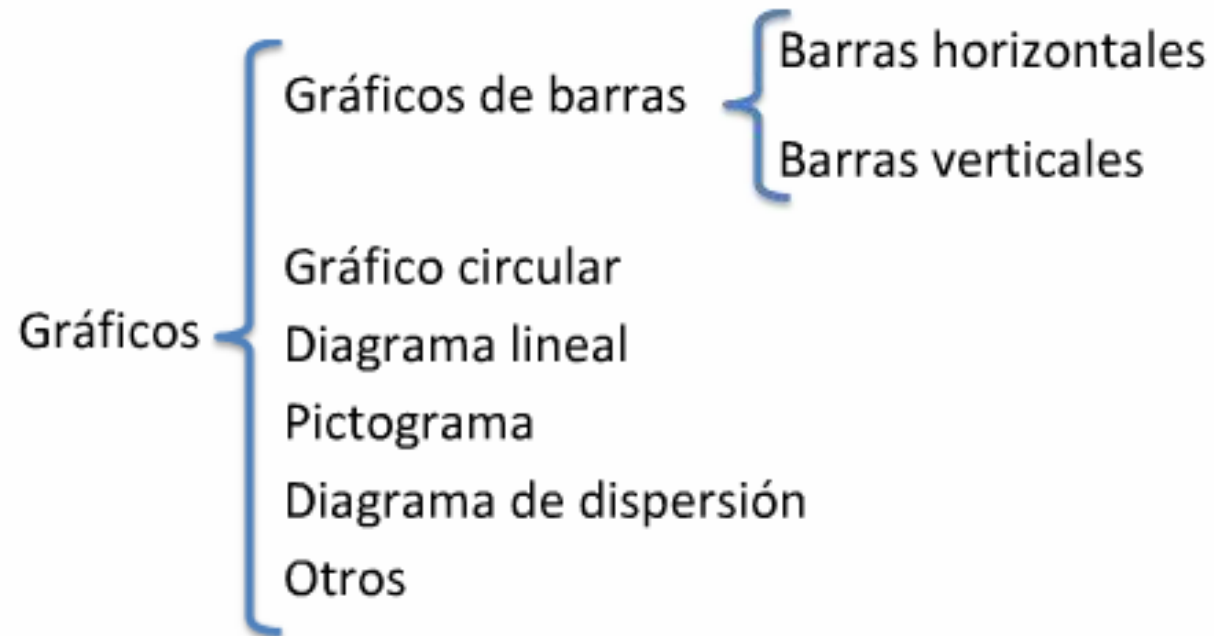
Presentación de los datos

Presentación gráfica



Presentación de los datos

Clases de gráficos



Presentación de los datos

Gráfica de barras horizontales simples

Se emplean para representar series cualitativas,
por ejemplo, series geográficas.

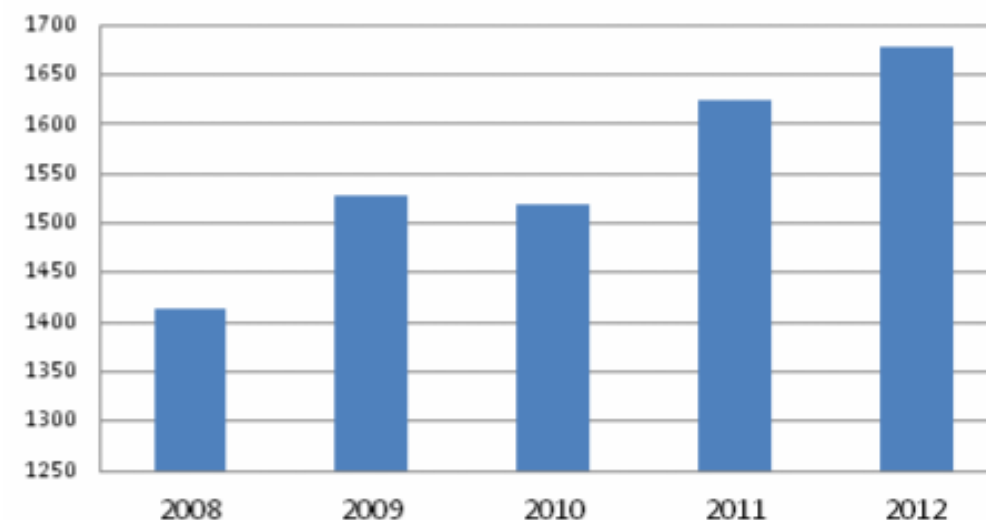


Presentación de los datos

Gráfica de barras verticales simples

Se emplean para representar series cuantitativas,
por ejemplo, series cronológicas.

República de Santa Clara: Exportaciones anuales 2008-2012
(En millones de dólares)



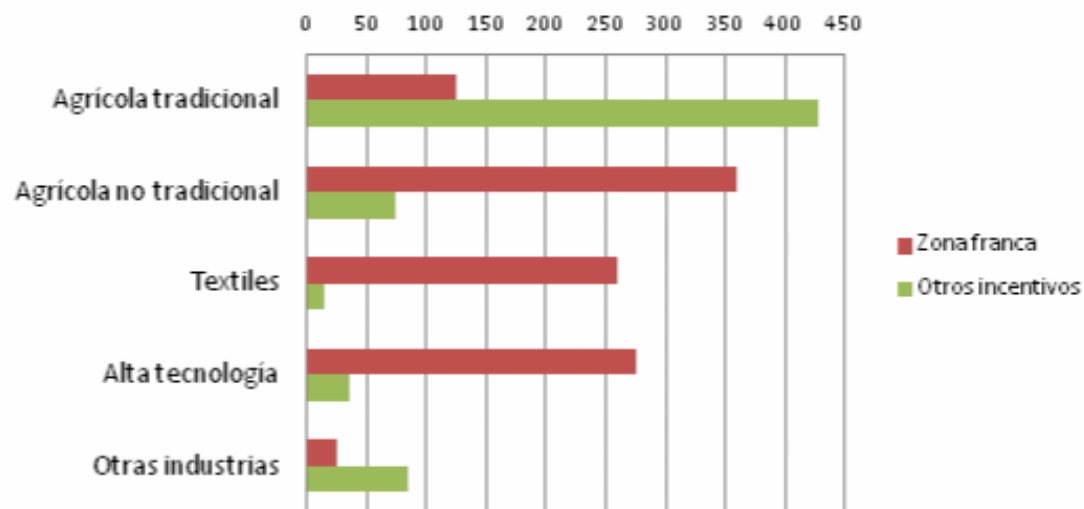
Presentación de los datos

Gráfica de barras horizontales comparativas

Se emplean para comparar varias series cualitativas.

**República de Santa Clara: Exportaciones por tipo de producto
y según sistema de incentivos
2012**

(En millones de dólares)

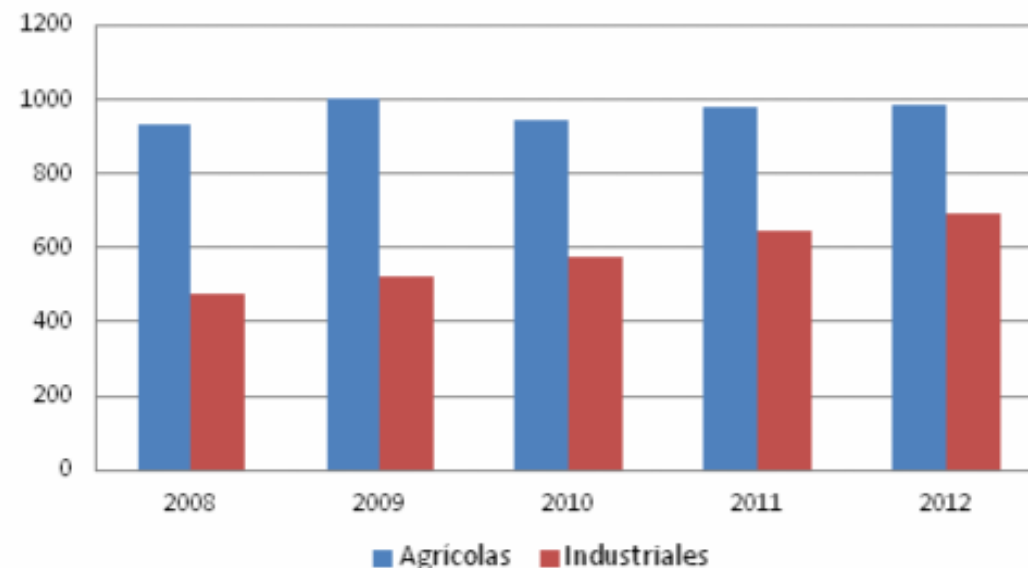


Presentación de los datos

Gráfica de barras verticales comparativas

Se emplean para comparar varias series cuantitativas.

**República de Santa Clara: Exportaciones anuales
por tipo de producto, 2008-2012**
(En millones de dólares)

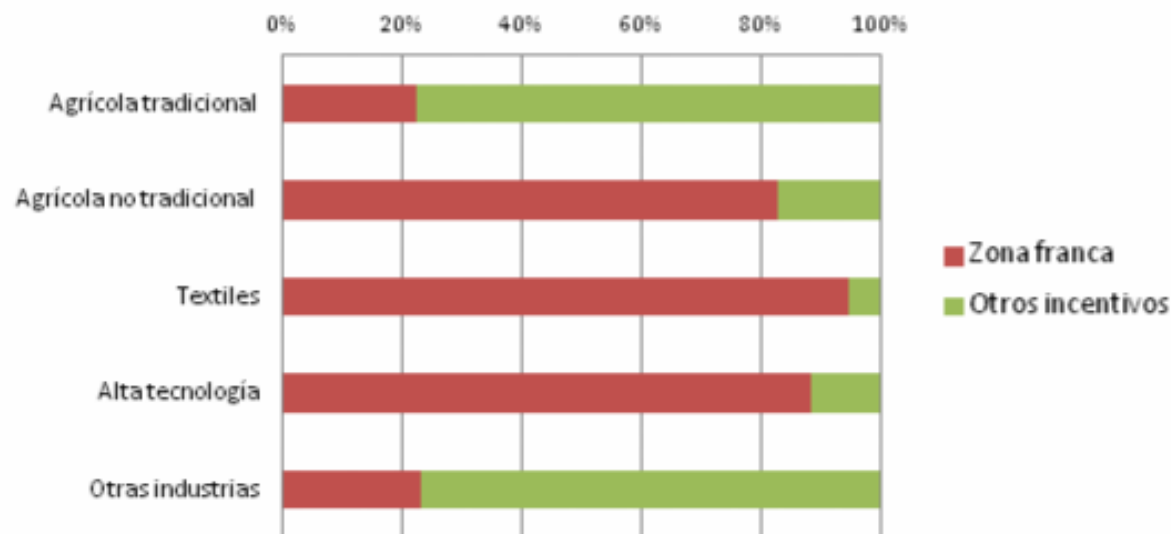


Presentación de los datos

Gráfica de barras horizontales compuestas

Se emplean para mostrar la composición de varias series cualitativas.

República de Santa Clara: Composición por sistema de incentivos según tipo de producto 2012

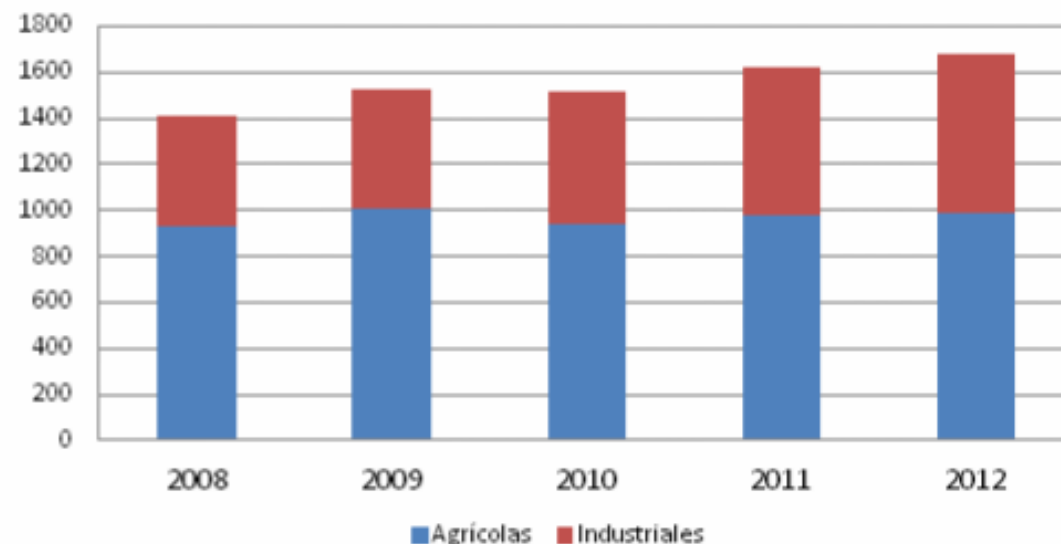


Presentación de los datos

Gráfica de barras verticales compuestas

Se emplean para mostrar la composición de series cuantitativas.

República de Santa Clara: Composición de las exportaciones anuales por tipo de producto 2008-2012
(En millones de dólares)

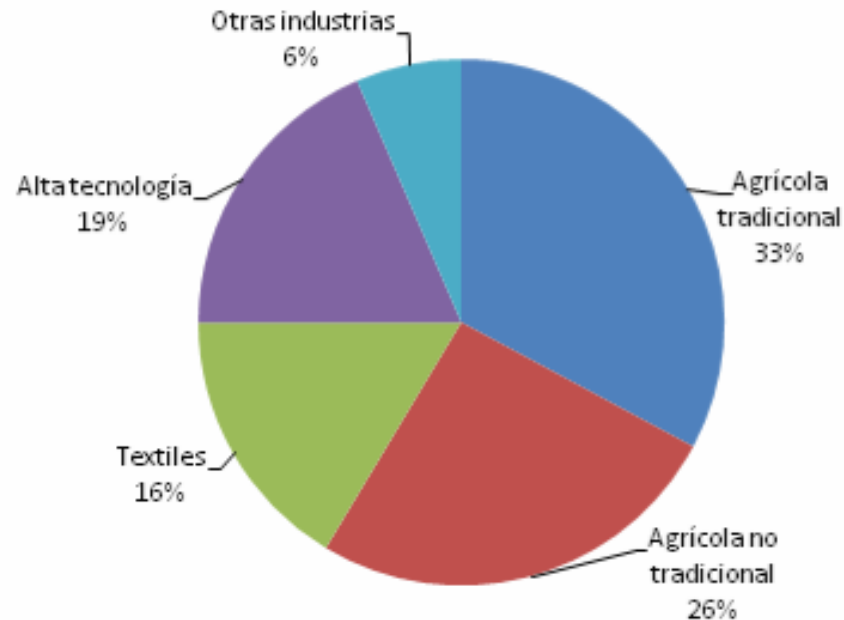


Presentación de los datos

Gráfica circular

Se emplea para mostrar la composición de un todo.

República de Santa Clara: Composición de las exportaciones por tipo de producto, 2012

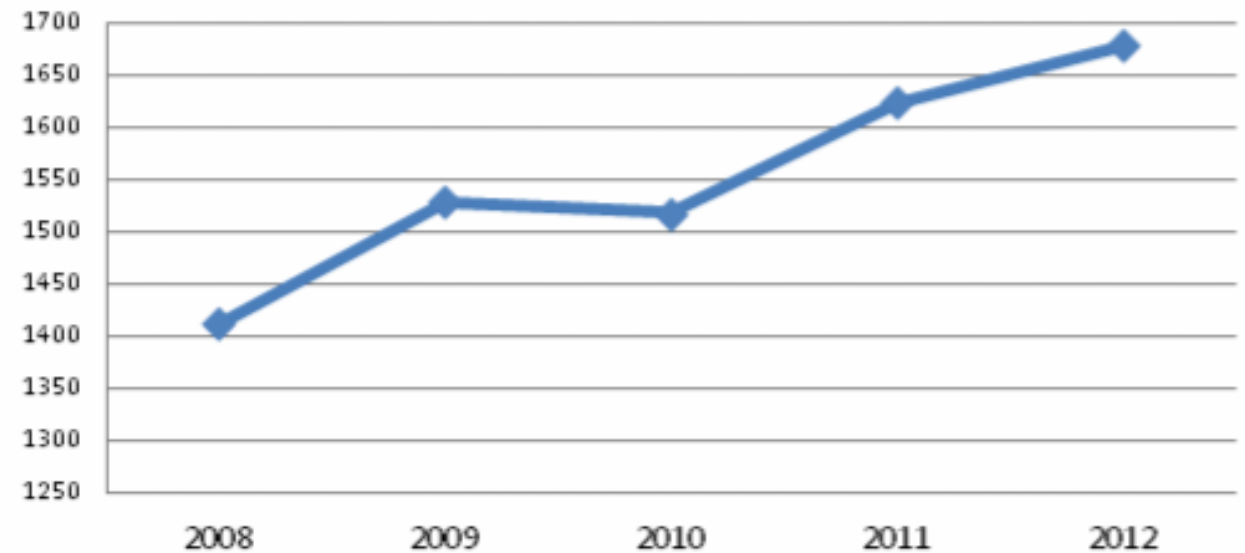


Presentación de los datos

Gráfica lineal

Se emplean para representar series cuantitativas, por ejemplo, series cronológicas.

República de Santa Clara: Exportaciones anuales 2008-2012
(En millones de dólares)

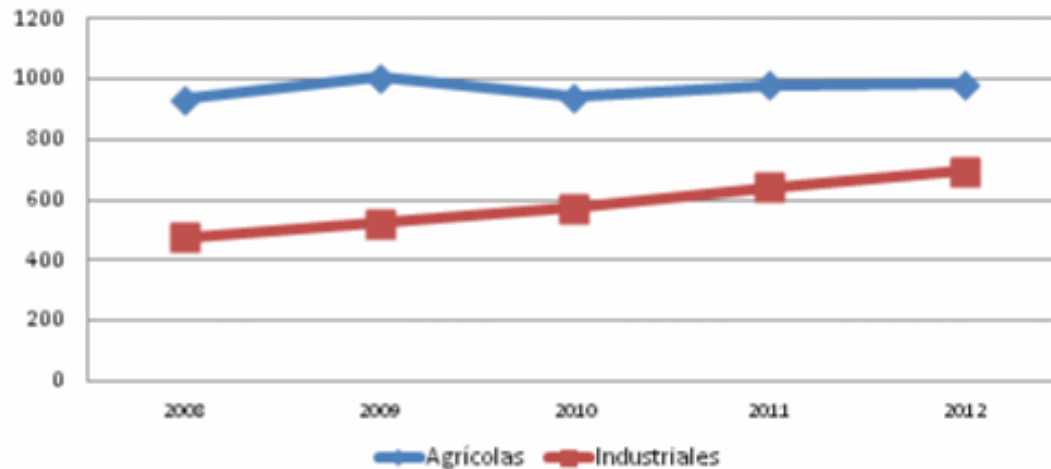


Presentación de los datos

Gráfica lineal

Se emplean para representar series cuantitativas, por ejemplo, series cronológicas.

República de Santa Clara: Exportaciones anuales por tipo de producto 2008-2012
(En millones de dólares)

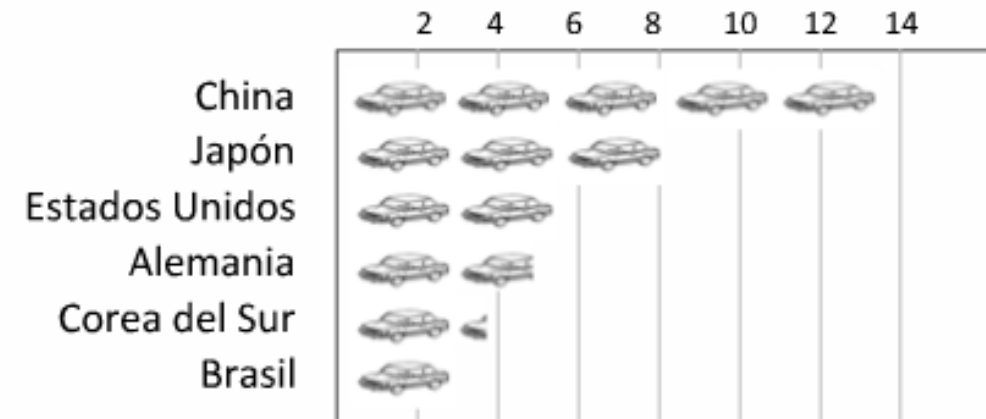


Presentación de los datos

Pictogramas

Utiliza imágenes o símbolos para representar los datos.

Principales productores mundiales de automóviles por países (en millones de unidades) 2001

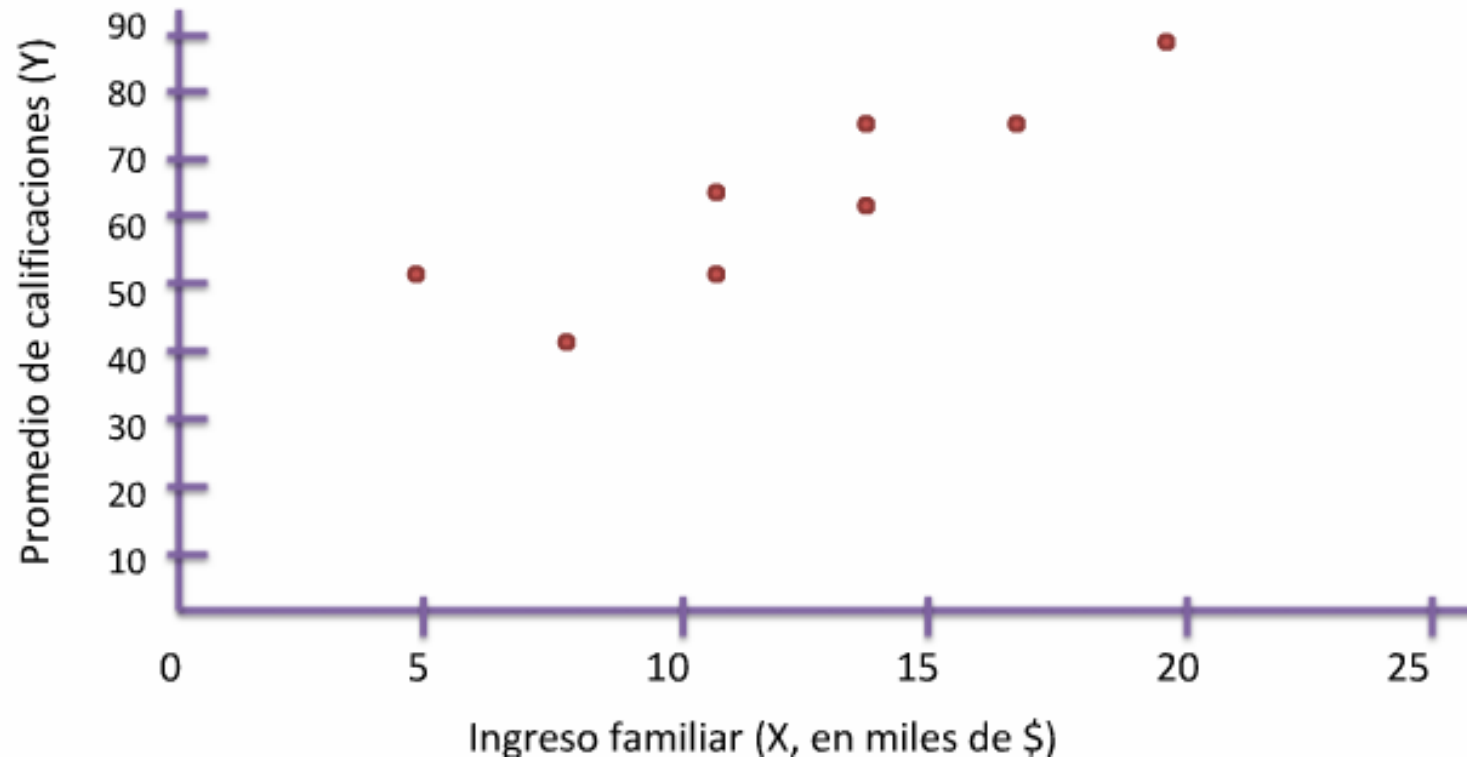


Fuente: PricewaterhouseCoopers

Presentación de los datos

Diagrama de dispersión

Se emplea para mostrar la relación entre dos variables.



Bibliografía

- ▣ Rodríguez Franco, Jesús, Pierdant Rodríguez, Alberto Isaac. Estadística para administración. (Primera Edición). México: Grupo Editorial Patria. (2014).
- ▣ Leandro Oviedo Gabriel. Estadística y Probabilidad con aplicaciones. (Primera Edición). Costa Rica: Publitex Grupo Editorial S.A. (2014).