Logo Ministerio de educación



**MATEMÁTICAS**

Guía de Apoyo Educativo en el área de las matemáticas.

Conceptos de la estadística y la presentación de datos en tablas y manera gráfica en estudios particulares.

Autor:

Adriana Quintero Palomino

TABLA DE CONTENIDO

[TABLA DE CONTENIDO 2](#_Toc437228791)

[TEMA 1: ASPECTOS GENERALES 7](#_Toc437228792)

[1.1 DEFINICIÓN DE ESTADÍSTICA 7](#_Toc437228793)

[1.2 TÉRMINOS QUE DEBEN CONOCERSE Y MANEJARSE 7](#_Toc437228794)

[1.2.1 Estadística descriptiva o deductiva 9](#_Toc437228795)

[1.2.2 Estadística inductiva o inferencial 10](#_Toc437228796)

[1.2.3 Estadístico 11](#_Toc437228797)

[1.2.4 Estadísticas 11](#_Toc437228798)

[1.2.5 Población 12](#_Toc437228799)

[1.2.6 Unidad o elemento 13](#_Toc437228800)

[1.2.7 Sustitución de elementos 13](#_Toc437228801)

[1.2.8 Característica 14](#_Toc437228802)

[1.2.9 Investigación total 16](#_Toc437228803)

[1.2.10 Investigación por muestreo 18](#_Toc437228804)

[1.2.11 Muestra 18](#_Toc437228805)

[1.2.12 Muestreo 20](#_Toc437228806)

[1.2.13 Parámetros 23](#_Toc437228807)

[1.3 MAL USO DE LA ESTADÍSTICA 34](#_Toc437228808)

[1.3.1 Fuentes de Información 35](#_Toc437228809)

[1.3.2 Clases de investigación 37](#_Toc437228810)

[1.4 ETAPAS EN UNA INVESTIGACIÓN 39](#_Toc437228811)

[Etapa 1. Planteamiento o formulación del problema 39](#_Toc437228812)

[Etapa 2. Formulación de objetivos 40](#_Toc437228813)

[Etapa 3. Definir la unidad de investigación 41](#_Toc437228814)

[Etapa 4. Formulación de hipótesis 41](#_Toc437228815)

[Etapa 5. Métodos de investigación 42](#_Toc437228816)

[Etapa 6. Recolección de datos 46](#_Toc437228817)

[Etapa 7. Organización y manejo de la información 46](#_Toc437228818)

[Etapa 8. Análisis de los datos 47](#_Toc437228819)

[Etapa 9. Redacción y entrega del informe final 47](#_Toc437228820)

[1.4.1 Selección del tamaño de la muestra 48](#_Toc437228821)

[TEMA 2: SUMATORIA 54](#_Toc437228822)

[2.1 Propiedades de la sumatoria 55](#_Toc437228823)

[Practica lo aprendido 59](#_Toc437228824)

[TEMA 3: DISTRIBUCIONES DE FRECUENCIAS 76](#_Toc437228825)

[3.1 PRESENTACIÓN DE DATOS 76](#_Toc437228826)

[3.2 CLASIFICACIÓN U ORDENAMIENTO DE LOS DATOS 76](#_Toc437228827)

[3.2.1 Formas de medición 77](#_Toc437228828)

[3.2.2 Cuadros 79](#_Toc437228829)

[3.3 TABLAS DE FRECUENCIAS 81](#_Toc437228830)

[3.3.1 Tablas o cuadros de frecuencias y representación gráfica 83](#_Toc437228831)

[3.4 VARIABLES 84](#_Toc437228832)

[3.4.1 Variable discreta 84](#_Toc437228833)

[3.4.2 Variable continua 91](#_Toc437228834)

[TEMA 4: GRÁFICAS CON APLICACIONES EN LA TEORÍA ESTADÍSTICA 99](#_Toc437228835)

[4.1 DIAGRAMAS DE FRECUENCIAS 99](#_Toc437228836)

[4.2 HISTOGRAMAS, POLÍGONOS Y OJIVAS 102](#_Toc437228837)

[Histograma de frecuencias 102](#_Toc437228838)

[Polígono 103](#_Toc437228839)

[Ojiva 104](#_Toc437228840)

[4.3 GRÁFICAS UTILIZADAS EN LOS INFORMES 106](#_Toc437228841)

[Gráfica circular o pastel 107](#_Toc437228842)

[Gráficas de barras o columnas 109](#_Toc437228843)

[Diagrama de una escala 110](#_Toc437228844)

[Gráficas de líneas 111](#_Toc437228845)

[Pirámide de población 114](#_Toc437228846)

[Pictogramas 116](#_Toc437228847)

[Cartogramas 117](#_Toc437228848)

[Gráficas de Gantt 119](#_Toc437228849)

[Diagrama de dispersión 120](#_Toc437228850)

[Gráfica en forma de anillo 121](#_Toc437228851)

[Gráfica en forma de radial 122](#_Toc437228852)

[Eterogramas 125](#_Toc437228853)

[Gráficas que utilizan cuadrados y triángulos 125](#_Toc437228854)

[4.4 ALGUNOS COMENTARIOS ACERCA DE LAS GRÁFICAS 127](#_Toc437228855)

[Practica lo aprendido 128](#_Toc437228856)

[BIBLIOGRAFÍA 140](#_Toc437228857)

**TABLA DE GRÁFICAS**

[Gráfica 1: Diagrama de frecuencia Xi 100](#_Toc437202050)

[**Gráfica 2:** Diagrama de frecuencias absolutas acumuladas (variable discreta) 100](#_Toc437202051)

[**Gráfica 3:** Histograma de frecuencias 102](#_Toc437202052)

[Gráfica 4: Curva normal 103](#_Toc437202053)

[Gráfica 5: Polígono de frecuencias Xi 104](#_Toc437202054)

[Gráfica 6: Ojiva de Frecuencias. 105](#_Toc437202055)

[Gráfica 7: Gráfico Circular de distribución de porcentajes por lugar de compra. 108](#_Toc437202056)

[Gráfica 8: Diagrama de una escala de Formas de PAGO. 111](#_Toc437202057)

[Gráfica 9: Representación Lineal 113](#_Toc437202058)

[Gráfica 10: Pirámide Poblacional. 115](#_Toc437202059)

[**Gráfica 11:** Pictogramas con el uso de figuras 117](#_Toc437202060)

[Gráfica 12: Cartograma de tasa de natalidad. 118](#_Toc437202061)

[**Gráfica 13:** Diagrama de dispersión 121](#_Toc437202062)

[Gráfica 14: Gráfica en forma de anillo 122](#_Toc437202063)

[**Gráfica 15:** Gráfico radial de relación de tipo de planta y ventas. 124](#_Toc437202064)

[**Gráfica 16:** Gráfico radial de relación de salarios. 124](#_Toc437202065)

[Gráfica 17: Forma de triángulo. 126](#_Toc437202066)

[Gráfica 18: Cubo de colores. 126](#_Toc437202067)

**TABLA DE TABLAS**

[**Tabla 1:** Número de fumadores de acuerdo con la marca preferida 82](#_Toc437202068)

[Tabla 2: Número de fumadores de acuerdo con las marcas que compra generalmente 83](#_Toc437202069)

[Tabla 3: Clasificación de empleados de una empresa, según estado civil y sexo 83](#_Toc437202070)

[Tabla 4: Datos originales o no agrupados 87](#_Toc437202071)

[Tabla 5: Tabulación. Variable discreta 88](#_Toc437202072)

[Tabla 6: Tabla de frecuencias. Variable discreta 89](#_Toc437202073)

[Tabla 7: Simbología. Tabla de frecuencias 90](#_Toc437202074)

[Tabla 8: Datos no agrupados. Variable continua 93](#_Toc437202075)

[Tabla 9: Tabulación. Variable continua 96](#_Toc437202076)

[Tabla 10: Simbología. Variable continua 97](#_Toc437202077)

[Tabla 11: Distribución porcentual sobre lugares de compra de 400 familias 107](#_Toc437202078)

[Tabla 12: Valores (Miles de Millones), en las formas de pago en 10 establecimientos comerciales durante un fin de semana en la ciudad x. 110](#_Toc437202079)

[Tabla 13: Evolución de los precios de tres productos del país: 2007 – 2011 112](#_Toc437202080)

[Tabla 14: Matrícula por sexo y cursos 114](#_Toc437202081)

[Tabla 15: Datos para publicación. 129](#_Toc437202082)

[Tabla 16: Cursos por trabajador. 129](#_Toc437202083)

[Tabla 17: Deudores morosos por sucursal. 134](#_Toc437202084)

[Tabla 18: Rango de Utilidades y Porcentajes de Utilidades 136](#_Toc437202085)

[Tabla 19: Votación. 138](#_Toc437202086)

[Tabla 20: Número de familias por marca. 139](#_Toc437202087)

# TEMA 1: ASPECTOS GENERALES

## DEFINICIÓN DE ESTADÍSTICA

No se ha logrado un acuerdo en torno al origen etimológico de la palabra **estadística.** Se­gún algunos autores, se deriva del latín **status** que significa estado, en el sentido político. Si se tiene en cuenta que en un principio los gobiernos la utilizaban para la recolección de datos, también significa "posición" o "situación de personas o cosas". Según otros, se deriva del vocablo griego **statera** que significa "balanza", en razón de la función que cumple en la medición o peso de los hechos que son objeto de estudio; otros sostienen que proviene del alemán **staat** que también significa "estado", considerado como una organización política.

No hay unidad acerca del origen de la palabra estadística; sin embargo, puede afir­marse que la etimología latina (**status**) es la más aceptada, por ser la que más explica la finalidad y aplicación de esta disciplina.

## TÉRMINOS QUE DEBEN CONOCERSE Y MANEJARSE

La palabra **estadística** se menciona y se utiliza a diario, con una gran variedad de inter­pretaciones. Sin embargo, la relacionan con todos o algunos de los siguientes términos**: recolección, clasificación, tabulación, descripción e interpretación de resultados, estimativos, proyecciones y adopción de decisiones frente a la incertidumbre.**

Se podría afirmar que la palabra estadística está asociada a términos o frases como: **análisis de datos; diseños de experimentos; procesos aleatorios; métodos cuantitativos; prue­ba de hipótesis; decisiones prácticas**. Por otra parte, en numerosas ocasiones cuando se dispone de muchos datos, el problema principal es cómo organizarlos, resumirlos, presentarlos e interpretarlos.

La estadística es una herramienta para resolverlo. A pe­sar de la buena impresión que a primera vista ofrece la estadística, también goza de impopularidad, especialmente cuando se hacen encuestas sobre simpatizantes por un candidato o por el gobernante del momento; lo mismo sucede con los indicadores económicos, entre ellos el índice de precios del consumidor, utilizado para indicar las variaciones en los precios de los artículos de primera necesidad, generalmente, aplica­do para calcular el porcentaje de inflación o pérdida del poder de compra del dinero respecto a un período determinado.

Para la mayoría de tratadistas, la estadística es un método de investigación, no solo para la observación y descripción de un hecho, sino también para la formulación de hipótesis, la aplicación del diseño de experimentos, la comparación de determinadas teorías, cuando no se puede hacer experimentalmente.

Otros la definen, a menudo, como un conjunto de métodos para la toma de decisio­nes en condiciones de incertidumbre.

Algunos la sintetizan, como el profesor Ross, quien da una definición que abarca a muchas otras, al afirmar que la estadística "**es la ciencia que estudia los hechos colectivos, mediante la observación numérica, el análisis matemático y la interpretación lógica, inves­tigando especialmente sus causas y leyes empíricas**".

Una definición propia, que sintetizaría gran parte de las definiciones existentes sería:

**La estadística es un conjunto de métodos aplicados en la recolección, clasificación, presentación e interpretación de datos numéricos, obtenidos de la observación de hechos que permiten, a su vez, inferir conclusiones lógicas.**

Vale la pena recalcar que la estadística moderna es la teoría de la información que como objetivo la inferencia.

**Términos y expresiones que deben conocerse**

Por lo anterior, el conjunto de métodos que ofrece la estadística cumple dos funciones igualmente importantes: **describir e inferir**, permitiendo su clasificación en estadística o deductiva y estadística inductiva o inferencial.

### Estadística descriptiva o deductiva

Como su nombre lo indica, describe el comportamiento de un hecho, a través de la recolección, el ordenamiento, clasificación, presentación de los datos mediante cuadros y gráficas, análisis e interpretación y utilización de medidas, algunas de ellas como los promedios.

La estadística **descriptiva o deductiva** podría resumirse como el conjunto de instrumentos que colaboran en la depuración de los datos, en su presentación clara y en su simplificación para la interpretación, el análisis y su publicación. A esto se le podría complementar, elaborando un cuadro que sea claro, acompañado de una buena gráfica que visualice mejor la información; además, la aplicación de una o varias medidas bien empleadas, que permita describir la "verdad" que encierra un conjunto de datos.

Algunos consideran que si solo se investigara la totalidad de las observaciones de una **población objetivo**, bastaría la aplicación de la **estadística descriptiva**, cuyo proceso sería, como ya se anotó, la recolección, presentación y análisis de la información obtenida. Se sabe, además, que en la mayoría de los casos eso no es posible por varios motivos, entre otros, por **costo y tiempo**, que obligan a realizar investigaciones más pequeñas denominas **muestras**, a través de las cuales se estiman los valores de la población, cuando el investigador requiere determinar la extensión, validez y confianza del proceso, valiéndose para ello de la **inferencia estadística.**

### Estadística inductiva o inferencial

Son todas aquellas técnicas usadas en la realización de investigaciones parciales, denominadas **muestras**, que permiten la obtención de ciertos resultados mediante la aplicación de medidas, considerados como **estimadores** de los valores estadísticos de la población, también denominados **parámetros**.

La **estadística inductiva** parte de datos que han sido depurados y condensados, facilitan el análisis, el establecimiento de relaciones y correlaciones, la aplicación de mediciones especiales que permiten inferir y contrastar hipótesis, en aquellos campos donde no existen verificaciones físicas.

### Estadístico

Aquella persona que aplica métodos estadísticos y maneja cifras, que presenta información de hechos, recogidos a través de la observación o experimentación. Algunos, por costum­bre, utilizan inapropiadamente el término de **estadígrafo**, que realmente significa medida.

### Estadísticas

Son todas aquellas informaciones o cifras que aparecen en las publicaciones en forma de cuadros o gráficas, las cuales se han ordenado de forma sistemática, para mostrar el comportamiento de un hecho que ha sido objeto de estudio.

### Población

En la forma más simple, puede afirmarse que este término se refiere a un conjunto finito o infinito de elementos. Es sinónimo de agregado y, cuando se emplea, no hace referencia únicamente a un conjunto de seres vivientes, a la población de una región o una ciudad, como se podría entender en principio, sino a un conjunto de observaciones totales.

El término **población**, también conocido como **universo**, se refiere a la totalidad o conjunto de medidas, o al recuento de todos los elementos que presentan una caracte­rística común. Se aconseja, con el fin de evitar equívocos, el término **población** (**objetivo**) **investigada** o por **investigar**.

De acuerdo con el número de elementos que constituye o conforma la población objetivo, se clasifica en **finita** o **infinita**. En el primer caso, se tendrá un número limitado de elementos o unidades; en el segundo, deberá ser ilimitado. Definir la cobertura, es decir, identificar y determinar el tamaño de la población, va a ser objeto de estudio más adelante, siendo de gran importancia su conocimiento en las primeras etapas de la investigación estadística.

Un conjunto de estudiantes matriculados en un plantel podría considerarse una población finita; pero si se hace referencia a todos los estudiantes de un país, no solo en el presente, sino en el pasado, se tendría una población infinita. En una industria manufacturera, serían numerosas las poblaciones de estudio, de acuerdo con el obje­tivo de investigación y, por consiguiente, con la variable principal que sea de interés para el investigador. Algunas de ellas podrían considerar características como: lotes o unidades producidas durante un periodo; cuentas por cobrar o de difícil cobro; nú­mero de empleados que laboran, etcétera.

### Unidad o elemento

Es el objeto, institución, región, persona, familia, animal, etcétera, que va a ser sometido a estudio, respecto a aquellas características que más interesan. Esta **unidad o elemento** debe definirse claramente, entenderse por todos, sin equívocos; además, debe permitir medirse o contarse, según al tipo de estudio y, finalmente, hacer comparaciones. El **ele­mento** es indivisible, en cambio, la **unidad** es divisible, es decir, se puede descomponer en otras unidades o en otros elementos.

### Sustitución de elementos

En una investigación por muestreo, cuando al realizar la encuesta, algunos de ellos no responden o no dan información por diferentes motivos, por tanto, no se debe o no se aconseja realizar sustitución del informante; sin embargo, en la práctica, por lo general, se hace de diferentes maneras:

1. Seleccionando un número mayor al tamaño de la muestra requerida.
2. Seleccionando al siguiente elemento que se consideró elegible dentro de la muestra.
3. Seleccionando el elemento que no informó dentro de aquellos que informaron, duplicando la información.
4. Dentro de la población que no quedó seleccionada, se puede realizar una nueva  
   selección, considerando un número igual al que no informó.

### Característica

**Los** elementos que conforman la población objetivo deben poseer ciertos rasgos o cualidades denominadas **características** que deben estudiarse. Por tanto, todos los elementos o unidades de esa población deben poseerlas, es decir, deben ser comunes en su presentación para que guarden relación con el **objetivo de la investigación**.

En los ejemplos de población, se mencionaron los estudiantes de un plantel; estos, como conjunto, constituyen la población objeto; el estudiante es el elemento y las características que se van a estudiar podrían ser: calificaciones, asistencia, edad, sexo, nivel o grado, procedencia, aficiones, entre otras. En el caso de una industria, las poblaciones serían el total de unidades o lotes producidos en un periodo; las cuentas por cobrar, el número de trabajadores, etcétera. Los elementos son: unidades, cuentas y empleados de la empresa, respectivamente, y las características que se van a estudiar, entre otras, serían: calidad, peso, resistencia, dureza del artículo; valor de las cuentas, tiempo de vencimiento, concepto de la deuda; horas trabajadas; accidentes de trabajo, asistencia, antigüedad, salarios devengados, cargos, estudios realizados, tiempos de experiencia, visitas al médico, tiempo de incapacidades, etcétera.

Según las anteriores características, se conforman dos grupos: cualitativas y cuantitativas.

**Características cualitativas**: Todas ellas se expresan mediante **palabras**, tales como: sexo (masculino, femenino); grado (primero, segundo,..., decimoprimero...); aficiones (canto, lectura, deportes, teatro,...); calidad (defectuosos, no defectuosos); estudios (economía, administración,...); entre otros. Estas y otras más se consideran **características cualitativas**, y su cuantificación se obtiene mediante el recuento: cuántos hombres y mujeres hay; cuántos alumnos hay en cada grado, etcétera. Estas se conocen también como **atributos**, y pueden admitir varias formas de presentación, denominadas **modalidades o categorías.**

**Características cuantitativas:** El segundo grupo se conforma por aquellas características que se expresan en forma **numérica**, ya que requieren ser medidas, como: peso, tiempo, número de accidentes o visitas al médico. A las anteriores se les consideratambién como **características cuantitativas**, conocidas como **variables**, las que, a su vez, teóricamente se clasifican en **variables discretas** y **continuas**.

1. **Variables discretas**: son aquellas que admiten únicamente **valores enteros**: el núme­ro de accidentes o visitas al médico, son algunos ejemplos. Las **variables discretas** se cuantifican mediante el recuento.
2. **Variables continuas:** admiten fracciones; por ejemplo, peso: 30 libras y 10 onzas; tiempo de experiencia dada en años, meses y días, etcétera. En las **variables con­tinuas** se procede mediante el uso de la medición.

### Investigación total

Denominada también **censo o investigación** exhaustiva, se presenta cuando se investiga la totalidad de los elementos que constituyen la población objetivo.

Cuando se hace referencia a un **censo** de población, de viviendas, agropecuario o industrial, por ejemplo, nos está indicando una enumeración completa de unidades o elementos, ya sean personas, familias, viviendas, establecimientos industriales o explo­taciones agrícolas en un tiempo dado, con unas características específicas.

La investigación total, en muchos casos, es impracticable; por ejemplo:

1. Cuando el número de observaciones es demasiado grande o infinito, es imposi­ble investigar a la totalidad de los elementos o unidades.
2. Cuando se requiere demasiado tiempo en su observación o recolección, presenta el inconveniente de desactualizarse cuando van a utilizarse.
3. Cuando los costos son elevados, superando los recursos disponibles para su rea­lización.
4. En casos como en el control de calidad, el elemento sometido a prueba puede destruirse, impidiendo la investigación de su totalidad; a esto podrán añadirse algunas situaciones más: por ejemplo, cuando la empresa no cuenta con personal suficiente para su realización; o cuando existe dificultad en la recolección de la información.

### Investigación por muestreo

Esta expresión se utiliza frecuentemente cuando no es posible la realización de una investi­gación total. Este procedimiento es menos costoso, más rápido de realizarlo, requiere me­nos personal y, finalmente, puede proporcionar una mayor precisión en los resultados.

### Muestra

Es la parte de elementos o unidades seleccionadas de una población que contiene una característica igual a aquella que no quedó seleccionada.

Corresponde a una parte de la población o subconjunto de un conjunto de elemen­tos, que resulta de la aplicación de algún proceso estadístico, generalmente selección de­liberada, con el objetivo de investigar las características de los elementos de la población o del conjunto del cual proviene.

También hay una **observación parcial o muestra**, cuando se investiga una parte de los de una población. Algunos la denominan como **subpoblación**, sin ninguna diferenciación; otros, por el contrario, la consideran aquella parte seleccionada, cuyos elementos presentan una característica tan especial, que el resto de la población no la tiene. Esa última apreciación no corresponde a la definición de muestra y más biencon el concepto de **estrato**.

De acuerdo con la probabilidad que muestren los elementos por seleccionar, la muestra puede ser aleatoria o no aleatoria.

1. **Muestra aleatoria:** deberá ser representativa de la población, si las probabilidades de selección de los elementos son todas iguales, durante todo el proceso. Por ejemplo, el método de selección por sorteo, aplicado a los estudiantes en un salón de clases; para ello se elaboran fichas ya sea con el nombre o el código del alumno de manera que se identifique al elemento, todas se introducen en una bolsa, luego se revuelven y van extrayéndose de una en una de acuerdo con el número de alumnos que han de seleccionarse. Este proceso puede hacerse con reposición o sin esta, según el criterio del estadístico.
2. **Muestra no aleatoria**: cuando algunos de los elementos de la población tengan más probabilidad de presentación que otros. Por ejemplo, la mayoría de las encuestas de opinión, en las que los elementos se seleccionan en forma caprichosa y por conveniencia; por esta razón, sus resultados no merecen confianza.

Denominada también como muestra no probabilística suele ser más sencilla y ba­nca de obtener. Algunas de ellas son denominadas muestras ajuicio, muestras por tanteo y aquellas en las que los elementos se seleccionan de manera voluntaria. Este procedimiento no es confiable, y por lo general conduce a errores. Solamen-k ñeñe alguna validez cuando la característica tiende a ser homogénea.

1. **Muestra probabilística**: es aquella que se ha denominado aleatorio, dentro de la cual se consideran los siguientes métodos: el muestreo aleatorio simple, estrati­ficado, por conglomerados, áreas o geográfico, sistemático, por fases y la combinación de algunos de ellos que recibe el nombre de mixto.

### Muestreo

En la utilización de cualquiera de los procedimientos o métodos de muestreo previamente establecidos para obtener una muestra, no debe perderse el objetivo, que es el de **considerar mayor número de elementos con el menor costo posible.**

* **Muestreo aleatorio o probabilístico**

Los métodos del muestreo aleatorio pueden ser:

1. **Muestreo aleatorio simple**

Como su nombre lo indica, los elementos de la población deben tener igual probabilidad de ser seleccionados para formar parte de lamuestra. En la aplicación de este método, en primer lugar, debe elaborarse un listado completo, actualizado y correctamente identificado, de los elementos que forman par­te de la población, y que van a ser seleccionados por algún método al azar, como el sorteo, extrayendo de un recipiente fichas o papeletas, cada una de ellas identificando a la unidad o al elemento, procedimiento similar al del bingo. Además, en su lugar se podrá utilizar la calculadora, oprimiendo primero la tecla SHIFT y luego, donde apa­rece RANN°, despreciando el cero (O,) que aparece en pantalla al comienzo de la cifra; también se puede hacer lo anterior, mediante el uso de las tablas de números aleatorios, construidas específicamente para ello.

Este método de muestreo, generalmente, se aplica en aquellos casos en que la pobla­ción investigada no es muy grande, presenta poca variabilidad y los elementos, además, muestran una reducida dispersión geográfica, es decir, presentan una gran concentra­ción en el lugar en el que se realiza el estudio.

1. **Muestreo estratificado** Generalmente debe aplicarse cuando la población requiere ser dividida en subpoblaciones o estratos, por presentar su característica un alto grado de variabilidad, buscando, de esta manera, que cada uno de los estratos sea lo más homo­géneo posible, y un elemento dado no pueda pertenecer a más de un estrato. Una vez establecidos los estratos, se elige, en cada uno de ellos, el número indicado de elementos considerados en la muestra aleatoria.
2. **Muestreo sistemático** Se aplica en aquellos casos en que la característica investigada esté ordenada, ya sea alfabéticamente, por valor o cantidad o en forma cronológica. El procedimiento de aplicación es simple: se calcula un intervalo de selección, dividiendo el tamaño de la población por el número de elementos en la muestra. Al seleccionar al azar un número dentro de ese intervalo, este se constituye en punto de iniciación o primer elemento seleccionado, y luego va sumándosele el valor del intervalo calculado, para la selección de los demás elementos.

 = intervalo de selección

1. **Muestreo por conglomerados** Se emplea a menudo para reducir el costo de muestrear una población dispersa en un área geográfica extensa. También se utiliza cuando no es posible elaborar un marco con elementos finales o de análisis; por tanto, se toma la unidad, es decir, aquella que contiene más unidades o elementos y que constituye el determinado conglomerado. En este caso, se seleccionan al azar los conglomerados y de ellos los elementos que van a constituir la muestra.
2. **Muestreo por etapas** Cuando de un marco se extrae una muestra de conglomerados, para luego seleccionar de ella otra muestra, y así sucesivamente hasta llegar a la unidad o al elemento de análisis o final, se le denomina muestreo por etapas. Algunos lo denomi­nan muestreo por áreas o geográfico o conglomerados.

* **Muestreo no aleatorio**

En este método, denominado también no probabilístico, las unidades no se seleccionan al azar, sino en forma caprichosa, por conveniencia o con alguna intención; de ahí se de­duce que sus resultados no son confiables. Este método incluye el muestreo voluntario, el muestreo de a juicio y el muestreo por cuotas.

1. **Muestreo de voluntarios** Es aquel en el cual las unidades o elementos no se seleccio­nan, son ellos quienes deciden formar parte de la muestra, colaborando voluntariamen­te, ya sea diligenciando un cupón que aparece en alguna publicación o realizando una llamada para participar, suministrando la información solicitada.
2. **Muestreo a juicio** Llamado también intencional, puesto que las unidades se seleccio­nan a juicio del investigador, ya sea por conveniencia o algunas veces con intención. También se denominan muestras dirigidas, y frecuentemente se utilizan en los procesos de auditoría.
3. **Muestreo por cuotas** A cada persona encargada de recolectar la información se le asigna cierto número de unidades o elementos muéstrales, cuya selección depende de la decisión que tome el entrevistador, aunque, necesariamente, haya necesidad de estable­cer o dar ciertas indicaciones.

### Parámetros

Son medidas aplicadas a las características de las unidades de la población. Frecuente­mente interesa determinar un valor de la variable que represente o muestre el compor­tamiento, ya sea del total de la población. Algunas de ellas podrían ser:

1. **La media o promedio** Por ejemplo, promedio de horas trabajadas, el ingreso medio de los empleados, el rendimiento medio de algodón por hectárea, y muchos más.
2. **El total** Por ejemplo, el total de establecimientos comerciales en una ciudad; el total de granos recolectados en una cosecha, el total de faltas de asistencia de alumnos a clase durante un periodo, etcétera.
3. **La proporción** Por ejemplo, la proporción de niñas que estudian bachillerato co­mercial; de establecimientos dedicados a la distribución de determinado artículo; la de bachilleres que prestan servicio militar obligatorio, entre otros.
4. **La razón** Por ejemplo, la relación entre el gasto de dinero en alimentos y vivienda; la relación o razón de hombres y mujeres, como, cuando se dice que por cada hombre hay dos mujeres en un establecimiento escolar, etcétera.

En el desarrollo de este libro se verán otros promedios como: **mediana, media geomé­trica, media armónica centro recorrido, cuartiles, deciles y percentiles**, sin olvidar que exis­ten muchos más.

Se mencionó anteriormente que estas y otras medidas, aplicadas a las características de la población, se les denomina parámetros o valores estadísticos de la población; y al aplicarse, a las características de las unidades o de los elementos en una muestra, se les denominan estimadores; algunos a estos últimos los identifican como estadígrafos.

* **Estimador**

Pitman establece alguna diferencia entre estimador y estimación. El primero se considera como una norma de procedimiento, utilizada para estimar una constante perteneciente a una población; y la segunda, para los resultados o valores obtenidos. El estimador se puede clasificar en puntual y por intervalos.

1. **Puntual:** cuando hace referencia a un solo valor o punto calculado.
2. **Por intervalos:** cuando corresponde a dos puntos, dentro de los cuales se consi­dera estará contenido el parámetro, con cierto grado de confiabilidad, el cual ha sido establecido por el investigador.

* **Inferencia estadística**

Uno de los objetivos de la estadística es obtener conclusiones o generalizar acerca del parámetro, mediante el cálculo de estimadores, en una muestra aleatoria extraída de una población.

* **Intervalo de confianza**

Indica los límites de precisión en una inferencia, es decir, al calcular una medida en la mues­tra, por ejemplo la media aritmética, esta podrá ser igual o diferente al del parámetro; por tanto, es preferible establecer unos límites dentro de la cual se pueda encontrar esta última con cierto grado de confiabilidad preestablecido, no mayor al 10%.

* **Error, sesgo o bias**

La diferencia que puede haber entre estimador y el parámetro se denomina errores o sesgo generalmente es cuantificado por el investigador. Hay también errores ajenos al muestreo, los cuales surgen de imperfecciones y equivocaciones en el registro, conteo y análisis de las informaciones; son fallas instrumentales y/o humanas.

* **Dominio de estudio**

Se produce cuando se establece el tamaño óptimo de la muestra, es decir, se cuenta con el mínimo de elementos que debe tener la muestra y, luego, se trabaja con un número inferior a él. También cuando el diseño de la muestra no corresponde a lo investigado;se selecciona un número de estudiantes y se analiza únicamente a aquellos que trabajan; lo lógico sería seleccionar un determinado número de alumnos, dentro del total que desarrollan alguna labor o trabajan.

* **Importancia y aplicación de la estadística**

La importancia que tiene la estadística radica en sus grandes aplicaciones en las diferentes actividades. En la vida cotidiana, se leen, interpretan o usan datos, muchas veces sin tener conocimiento de lo que es la estadística, pero una persona con algunos principios estadísticos sacará mejor provecho de la información. Si se trata de un profesional, la importancia de estos instrumentos será mejor, ya que se utilizará en su campo de acción.

La estadística se considera una herramienta de vital importancia en la toma de decisiones, pues suministra la información necesaria para la valoración de planes y programas; además, su adecuada utilización dependerá en especial del conocimiento, habilidad y experiencia de quien la emplea.

Vale la pena repetir lo anotado por Salomón Fabricant, sobre su importancia: "Todo el parece coincidir en que la estadística puede ser útil para comprender, evaluar y controlar el funcionamiento de la sociedad". Por otra parte, H. G. Wells decía:

"El pensamiento estadístico un día será tan necesario para una ciudadanía eficiente como el saber leer y escribir".

Todo lo anterior se corrobora, si se observa que la mayor parte de los acontecimientos que diariamente ocurren en la sociedad, en cualquier campo y parte del mundo se encuentran estrechamente ligados a la concepción de número y medida.

Se podría afirmar, sin temor a equivocación, que la mayor parte de las actividades del ser humano o de las instituciones, por no decir que todas, no importa la rama en que se desarrollen, están afectadas en mayor o menor grado por decisiones basadas en antece­dentes de tipo cuantitativo.

Se podría comprender la importancia que tiene la estadística, si se observan desprevenidamente algunos de los siguientes puntos de reflexión:

1. Los métodos estadísticos se utilizan a diario, tanto en el sector público como en el privado.
2. Las noticias, frecuentemente dadas en los diferentes medios de comunicación, requieren interpretación estadística.
3. Algunas de las materias del plan de estudios necesitan la aplicación de estos co­nocimientos.
4. Para conocer su rendimiento académico a través del promedio de calificaciones ola posición ocupada en el curso, se utiliza desprevenidamente la estadística.

Es tan vasto el campo de aplicación que se le ha dado a la estadística, que sería larga su enumeración y podría resumirse diciendo que todas las ciencias, de una u otra forma, la aplican. Sin embargo, podría afirmarse que en todas las actividades: económicas, el empresario, el industrial o el hombre de negocios, en general, la utilización para el análisis financiero, en las ventas o en los procesos de la producción; en la investigación de mercados, para el lanzamiento de un producto, para averiguar su grado de aceptación o para la proyección de su producción y de sus ven­tas; cuando se desea aceptar o rechazar un lote de su producción o de la materia prima comprada; en la elaboración del presupuesto, o como parte integrante en el desarrollo de políticas y decisiones de carácter administrativo.

El Estado la utiliza frecuentemente para trazar políticas económicas y administrati­vas; en la elaboración de planes de desarrollo, ya sea para todo el país, para una región o un sector especial; en la elaboración del presupuesto de inversión y de funcionamiento; en la formación de políticas, ya sea en lo fiscal, aduanero, carcelario, educativo, etcétera.

También se utiliza en las campañas electorales, para determinar el grado de aceptación o popularidad que presenta el candidato a una corporación pública.

En la salud y la medicina, el Estado, los fabricantes de drogas, las instituciones o profesionales en el campo de la salud, realizan investigaciones encaminadas al trata­miento de ciertas enfermedades, como el sida, o al mejoramiento de los productos existentes, buscando que sean más efectivos; o para determinar los posibles efectos secundarios que puedan tener; también en el desarrollo de programas sanitarios, en la prevención de enfermedades.

En el sector agrícola, por ejemplo, se observa la aplicación de la teoría mendeliana de la herencia, en la obtención de nuevas variedades de mayor rendimiento o más re­sistentes frente a plagas o enfermedades; en la aplicación de fertilizantes e insecticidas, etcétera. En este campo se observa, así como en la biología, la utilización de métodos estadísticos para la realización de investigaciones que permitan detectar las reacciones no solo de las plantas sino de los animales.

Hoy día es preocupación permanente el aumento y mejoramiento de los recursos alimenticios, para atender a una población cre­ciente y, en gran parte, en condiciones nutricionales nada satisfactorias, lo cual requiere el apoyo de los métodos estadísticos.

En física y química se emplea con frecuencia, y una de sus aplicaciones es la contratación de hipótesis con base en resultados obtenidos a través de experimentos, aplicando la técnica estadística conocida como diseño experimental.

La estadística tiene un vasto campo de aplicación, tanto en las empresas comerciales como en las industriales. Si se considera que la empresa es una organización dentro de la cual se realiza una serie de operaciones, casi siempre, con la meta de obtener benefi­cios del capital invertido, se requiere información organizada que sirva de control y que permita hacer planes futuristas.

La estadística, dentro de la empresa, cumple, entre otras, las siguientes funciones:

1. Permite obtener una visión general de la empresa y de algunos de sus aspectos en particular.
2. Facilita descubrir las relaciones causa efecto en las distintas manifestaciones eco­nómicas.
3. Ayuda a analizar y a separar lo que es normal de lo que no lo es, es decir, permite ejercer control en muchas de las actividades u operaciones de la empresa.

La estadística aplicada en una empresa es importante para: la planeación de operacio­nes, el control de calidad del producto, la determinación del mercado, en las actividad financieras, en el control y elaboración del presupuesto, en la realización de inventarios, en la administración de personal, etcétera.

Se podría seguir mencionando una gama de actividades, profesiones o campos don­de la aplicación de la estadística es necesaria para el economista, el contador, el auditor, el ingeniero en todas las ramas, el actuario, el sociólogo, el geólogo, el mercadotecnista, el periodista,... todos ellos tendrán que aplicar de alguna manera estas técnicas en su ejercicio profesional.

Hoy, el uso del computador y algunas calculadoras manuales facilitan la utilización de paquetes estadísticos, como: SAS, MICROSTAT, TSP, SPSS y otros más, necesarios para el manejo de abundantes datos numéricos, permitiendo su agilización y la aplicación de una variedad de medidas que estos paquetes ofrecen, que facilitan el proceso de análisis y la toma de decisiones.

En todos los campos, la utilización de los métodos estadísticos se encamina hacia:

1. Aquellos procesos que permitan recolectar, agrupar, concentrar o descubrir con mayor facilidad ese cúmulo de información, a través de lo que algunos llaman "reducción de datos".
2. El análisis de los datos, mediante la observación directa o el desarrollo de experi­mentos, con el fin de realizar estimativos o comprobación de hipótesis.
3. Determinar o indicar los cambios que sufren los datos a través del tiempo, o estimar su comportamiento futuro.
4. Determinar las causas que originan esos cambios y permitir relacionar dos o más  
   hechos.

Dentro de la múltiple utilidad que ofrece la estadística, vale la pena recordar que todas aquellas actividades que se realizan en procura de un mejoramiento económico, en salud y educación, elevando, por consiguiente, las condiciones de vida, están liga­das a los programas de desarrollo de un país a través de sus instituciones, las cuales necesitan detectar y diseñar políticas basadas en informaciones recogidas, tratadas y analizadas mediante la aplicación de métodos estadísticos. Sin embargo, muchas ve­ces se critica al gobierno de turno de no contar con información estadística confiable, sin saber que estos hechos de ocurrencia diaria o periódica requieren cumplir ciertas condiciones para que formen parte de las estadísticas. Estas se describen a continua­ción.

1. Deben manifestarse al exterior, es decir, no se trata únicamente de observar su presentación, como el contrabando, pues para ello bastaría visitar los "Sanandresitos" o lugares donde generalmente se realizan. La manifestación o su observa­ción debe hacerse mediante registros. Si los registros presentan deficiencias, estas se reflejarán en las estadísticas. Muchos usuarios critican las estadísticas del sector oficial así como las del sector privado, pero no se dan cuenta de la falta de con­ciencia o el desinterés de algunas personas hace que muchas de las informaciones no se registren; tal es el caso de los robos, violaciones domiciliarias, entre otras, donde la gente no acude a colocar el denuncio correspondiente. Así que los indi­cadores sobre estos hechos solo harán referencia a las informaciones registradas y solo sobre ellas se aplicarán los instrumentos de medición. Así, aquellos hechos que no se registran, no serán abarcados por la estadística.
2. Además de ser registrados, para su estudio requieren que correspondan a hechos colectivos, descartando los individuales; es decir, lo sucedido a una persona no será analizado en forma particular, debe, entonces, formar parte de un grupo, masa o colectivo de datos para su estudio. El registro de la producción de una empresa solo tendrá un interés particular, pero si el gobierno es quien piensa intervenir el mercado, la información formará parte del sector que se analiza.
3. Se podrá agregar otra condición, como la de su frecuente repetición. Un hecho que se presente en forma aislada, es decir, ocurre una vez y difícilmente vuelve a ocurrir, no es observado por la estadística y queda solo como un registro histó­rico. Cuando dice que el hecho debe repetirse, no se hace referencia a que ocu­rra a cada momento; es posible, en algunos casos, que transcurra cierto tiempo para que se repita, tal como sucede con aquellos productos estacionales, o en el caso de las matrículas que ocurren cada año o cada semestre. Debido a su presentación en forma periódica, se les considera como hechos distantes en el tiempo.
4. Además de su frecuente repetición, deben ser a la vez de distinta frecuencia. Si la frecuencia es constante, de este hecho no se ocupa la estadística. Por ejemplo, las transacciones comerciales, el volumen o valor de las ventas que se registran todos los días, los nacimientos, defunciones, etcétera, sí son de interés para la estadís­tica, pues la intensidad es diferente de un periodo a otro. Si, por el contrario, el valor o la cantidad fuera igual a cada momento, bastaría registrar una sola vez. No se requiere que el hecho se presente en el mismo lugar, puede estar ocurrien­do en diferentes sitios; por tanto, la estadística abarca también aquellos hechos distantes en el espacio.
5. Finalmente, se podrá decir que aquellos hechos cuyas características cualitativas pueden cuantificarse, son tenidos en cuenta por la estadística. El amor a la patria o al trabajo no se pueden medir o cuantificar, pero sí el número de personas por cargos, ocupación, sexo, profesión, el número de artículos vendidos por departa­mento, marcas, modelos, sucursales, ..., son ejemplos de características cualitati­vas que sí las abarca la estadística.

## MAL USO DE LA ESTADÍSTICA

Después de hacer alusión al campo tan amplio de aplicación y la importancia de la esta­dística en todos los aspectos de la vida diaria, puede concluirse que ella puede emplearse para probar o mostrar cualquier hecho que ocurra. Algunos medios de información, que parecen veraces, mediante las estadísticas demuestran lo que ellos creen que puede ser cierto o falso, información que muchas veces debe aceptarse sin discusión. Otras de las estadísticas correctamente diseñadas y elaboradas se consideren engañosas, por el desconocimiento que se tiene de las técnicas aplicadas en su elaboración.

La mala interpretación, en algunos casos, se debe a que no satisface intereses personales; en otras ocasiones, se presentan acomodadas a determinadas circunstancias, pero generalmente es la falta de información sobre la metodología empleada, es decir, cómo se diseñó, qué grado de confiabilidad tiene, cuál es su cobertura y cuál es su verdadero significado o intención, si lo tiene.

Mi preocupación en estos momentos es el enfoque de este libro, si se debe conservar la forma tradicional de presentación, tal como ocurre en la mayoría de los textos o si, por el contrario, nos concentramos en las etapas de planeación, recolección, sistematización, la forma adecuada de presentación de los datos (cuadros, gráficas) y análisis de los resultados obtenidos a través del uso de la calculadora, del Excel o cualquiera de los paquetes estadísticos, ya que estos se incluyen en la parte operativa, mediante la aplicación de fórmulas. En este último proceso, se pensaría que solo se debería hacer énfasis en enseñar qué medidas o procedimientos son los más adecuados y el porqué de su utilización óptima. Mi decisión, a esta altura de elaboración, creo en la combinación entre las dos formas de presentación de tal manera de poder alcanzar los objetivos propuestos desde su inicio.

### Fuentes de Información

Las fuentes de información se clasifican en: internas, externas, primarias y secundarias.

* **Fuentes internas**  Allí los datos se producen y se registran internamente dentro de la empresa o entidad, en donde se puede obtener información sobre ventas, producción, cuentas por pagar, consumo de materia prima, horas trabajadas, etcétera.

Quizás por desconocimiento, o ignorancia, muchos no le dan importancia a las fuentes de información interna, y a veces las consideran como operaciones corrientes que forman parte de la rutina de los departamentos de contabilidad, ventas, producción, recursos humanos, etcétera, pues desconocen que la verdadera investigación comienza es allí, para iniciar, de esta manera, la búsqueda de información externa, por no encontrarse esta dentro de los archivos y documentos de la empresa.

* **Fuentes externas** Son aquellas que suministran informaciones o datos de lo que sucede fuera de la empresa. Por ejemplo, una encuesta acerca de las personas que utilizan los productos de la empresa, a fin de conocer el grado de aceptación.
* **Fuentes primarias** Cuando una persona, entidad o empresa realiza una investigación, con las consiguientes etapas de recolección, procesamiento y análisis, esta información se constituye en una fuente primaria.

Algunos denominan a las fuentes primarias como métodos de observación directa, ya que la información la recoge el interesado mediante la aplicación de un formulario o cuestionario. Gracias a aquellas se observan o se registran los datos directamente de la fuente de información, es decir, de las unidades que son objeto de estudio.

* **Fuentes secundarias** En cambio, para otra persona, entidad o empresa diferente a quien la produce, la informa­ción se convierte en fuente secundaria. Estas se obtienen de boletines, revistas, periódicos, publicaciones especializadas y de los informes realizados por otras entidades o empresas.

### Clases de investigación

Se menciona a Inglaterra como el país donde se inician las investigaciones de mercado y de opinión pública; sin embargo, en América del Norte tomaron mayor auge, y cada día se ve en todos los países o lugares, cómo las entidades gubernamentales, y en especial las empresas privadas, cuentan con departamentos de investigación.

La investigación se podrá clasificar en dos clases: descriptiva y experimental.

* **Investigación descriptiva** Se realiza sobre hechos reales con el fin de explicar, de modo sistemático, sus caracte­rísticas. Esta clase de investigación comprende etapas como la recolección, descripción, análisis e interpretación de los datos; también puede referirse al análisis de datos prove­nientes de fuentes secundarias. Se puede afirmar que, generalmente, se utiliza el método de casos y el estadístico.

En el departamento de mercados de una empresa, la investigación descriptiva podría estar encaminada, por ejemplo, a describir las características del mercado de un pro­ducto o de la empresa en general; la aplicación de algunas medidas, para determinar la proporción de personas que prefieren esa marca, es decir, conocer la reacción de los consumidores frente al producto, o el comportamiento futuro de las ventas, etcétera.

* **Investigación experimental** Corresponde a una observación dirigida en unas condiciones rigurosamente controla­das, determinando las relaciones de causa y efecto. Debe quedar claro que el experimen­to es una situación provocada, siendo esencial el empleo de un diseño adecuado para evitar, de esta manera, situaciones que invaliden la investigación.

La investigación experimental difiere de la investigación descriptiva en que la primera tiene su base en realidades físicas, y la segunda, en abstracciones; además, si las condi­ciones no varían, no se puede actuar sobre ellas, y tampoco reproducirlas, como cuando se realiza una experiencia en el laboratorio para comprobar su desarrollo.

Algunos la denominan modelos metodológicos y la clasifican en: (a) histórica; (b) expe­rimental; (c) descriptiva; (d) correlacional; (e) ex post facto sobre hechos cumplidos; (f) cuasi experimental; (g) teórica; (h) participativa; (i) evaluativa; (j) estudios de casos; (k) bibliográfica.

## ETAPAS EN UNA INVESTIGACIÓN

El modelo descriptivo, que también se denomina estadístico, requiere para su ejecución nueve etapas, que se describen a continuación.

### Etapa 1. Planteamiento o formulación del problema

El conocimiento del problema plantea la necesidad de realizar una investigación. Por ejemplo, un empresario siempre estará atento a la situación interna de la empresa, lo cual se constituye en tarea diaria de dirección. Evalúa los resultados de gestión frente los objetivos trazados, que van visualizándose a través de informaciones recogidas en los diferentes departamentos, sobre: ventas, participación de mercado, costos, liquidez, imagen, etcétera, que en algunos casos podrían mostrar anormalidad, síntomas sobre los cuales se deben diseñar los respectivos correctivos por aplicarse, con base en la formulación de una serie de hipótesis.

La decisión de realizar una investigación nace cuando se siente la necesidad de diseñar correctivos o fijar políticas a corto o largo plazo, que mejoren la situación de la empresa, y cuando no se disponga de una información adecuada, lo cual requiere la elaboración de un plan de investigaciones que permita la recolección y el análisis de sus resultados.

Se debe tener muy en cuenta que las conclusiones estadísticas dependerán en gran parte de la información recogida y esta, a su vez, de la formulación del problema.

### Etapa 2. Formulación de objetivos

Claramente definidos, los objetivos indican aquello que se desea hacer.

* ¿Qué se va a investigar?
* ¿Cómo se va a realizar?
* ¿Cómo se pueden obtener los datos? ¿Mediante una investigación parcial o total?
* ¿Cuál sería el diseño de muestreo?
* ¿Qué cobertura tendría?
* ¿Dónde se va a investigar?
* ¿En qué lugar?
* ¿Qué áreas deberán incluirse?
* ¿Cuándo es el momento más indicado para llevarla a cabo?
* ¿Cuándo se necesitan?
* ¿Para qué se requiere la información?
* ¿A quién va dirigida?
* ¿Cuál es la unidad de investigación?

Algunos establecen diferencias en cuanto al elemento y la unidad. La primera es indivisible, por ejemplo una persona, un objeto; la segunda, podrá dividirse, por ejemplo, en barrios, estos en manzanas, estas en viviendas, luego en hogares y a su vez en perso­nas. Esta última corresponde al elemento.

La claridad del objetivo dará una visión perfecta de lo que se va a hacer, es decir: (a) el fin de la observación; (b) la elección y definición de la unidad; (c) la lista de los elementos denominado marco o marco de selección y de las características; (d) el momento en que se va a realizar; (e) el método de observación; (f) las condiciones en que se desarrollará; (g) la elección de las medidas que se emplearán.

### Etapa 3. Definir la unidad de investigación

La unidad de investigación se definirá cuando los objetivos de la investigación se hayan determinado con precisión. Se podrá saber si hay necesidad de otras unidades de selec­ción para llegar al elemento o la unidad final, es decir, el objeto de estudio. Lo anterior permite elaborar el marco o marcos, es decir, conformar y actualizar el listado de uni­dades o elementos de la población que se seleccionarán al azar o en forma sistemática. De esta manera, se podrá establecer con exactitud la cobertura, el tiempo necesario, la naturaleza cuantitativa y cualitativa del hecho. No se debe olvidar que las unidades deben ser accesibles al observador.

### Etapa 4. Formulación de hipótesis

Las hipótesis son supuestos que se hacen en una investigación para luego ser comproba­dos. Las hipótesis estadísticas son aquellos supuestos respecto a valores de la población o parámetros, y cuya comprobación requieren la aplicación de métodos estadísticos. Aquellas hipótesis que no hacen referencia a un valor (media, proporción, razón, etcé­tera), son simplemente hipótesis.

Algunos ejemplos de hipótesis descriptivas son: "El rendimiento del curso en el pre­sente periodo lectivo es bueno"; "En un gobierno de coalición, la inmoralidad de los funcionarios disminuirá"; "Si hay inversión social, la situación de orden público mejora­rá". Como se observa en los ejemplos anteriores, no se hace mención de valor alguno.

En cambio, "Los salarios en el año venidero aumentarán por encima del índice de precios al consumidor" o "Los precios en diciembre subirán en una proporción inferior a la del año pasado" o "El promedio de calificación del curso será superior a la califica­ción de 4.1", todos ellos, son ejemplos de hipótesis estadísticas.

### Etapa 5. Métodos de investigación

Ya se anotó que la investigación puede ser total (censo) o parcial (muestra), dependien­do de ciertos factores que también se mencionaron: costos, tamaño de la población, tiempo. Sabiendo, además, que cada elemento tendrá cierta cantidad de información, se deberá seleccionar la estrictamente necesaria, dependiendo también del número de unidades o elementos que se va a investigar, su dispersión geográfica y el grado de varia­bilidad que la característica presenta.

El diseño de la encuesta por muestreo ayudará a controlar la cantidad de infor­mación y su variación, y en la decisión sobre cuál sería el método más indicado en esa investigación. Se procederá, luego, a establecer el método de observación o re­colección de datos, siendo los más usuales, entre otros: las entrevistas personales, por teléfono, la observación directa, cuestionarios enviados por correo. Cada uno de ellos presentará ventajas y desventajas. Cualquiera de estos métodos que se vayan a aplicar, deberá tener muy en cuenta, al elaborar preguntas, minimizar la falta de respuestas o contestaciones incorrectas. Al diseñar un cuestionario, vale la pena tomar en cuenta lo siguiente:

1. Incluir únicamente aquellas preguntas que sean estrictamente necesarias, puesto que un elevado número de ellas incrementará su costo y tiempo.
2. Formular preguntas claras, concisas, utilizando términos a nivel del informante, para que sean entendidas por todas aquellas personas involucradas en el estudio.
3. No hacer preguntas que de antemano se está seguro de que no van a ser con­testadas.
4. Iniciar con las preguntas generales, para luego seguir con las específicas.

Existen tres (3) formas para hacer una encuesta: entrevista personal, entrevista telefónica y el uso del correo.

* **Entrevista personal** Es un método muy costoso e implica un gran esfuerzo de adiestramiento a fin de tener la seguridad de que el entrevistador podrá extraer con habilidad la información deseada. Las ventajas principales de la entrevista personal son: el porcentaje de respuestas que se logra hacer suele ser el máximo, y el entrevistador puede con más facilidad las ambigüedades imprevistas.
* **Entrevista telefónica** Al igual que la anterior, es un método muy costoso porque requiere amplia capacitación del encuestador; no obstante, los porcentajes de respuesta no suelen ser altos, porque no hay un contacto, entre el entrevistador y el entrevistado, directo.
* **Encuesta por correo** Aunque es la menos costosa, por lo general presenta un mínimo porcentaje de respuesta. En ocasiones, las personas que en un momento dado las contes­tarían, van posponiendo el retorno de los formularios y, por ende, producen demoras o, en algunos casos, faltas de respuesta a muchas de las preguntas, que en vez de permitir la conclusión del proyecto, lo retardan de forma considerable.

Las preguntas, hasta donde sea posible, deben ser cerradas, en las que el informante tiene dos posibilidades de respuestas: sí o no. También podrían ser de respuesta numérica simple; por ejemplo, la edad, para que una persona coloque en el espacio indicado, el número de años o para que elija uno de los intervalos preestablecidos. Se acostumbra, con mucha frecuencia, utilizar preguntas de selección múltiple, en los que la persona po­drá señalar más de una respuesta, como en la pregunta referente al deporte o deportes que practica. Las preguntas poco recomendables, por la dificultad que presentan en su tabulación, son las abiertas, que permiten la libre expresión de acuerdo con la opinión que se tenga en el momento de la entrevista.

Antes de aplicar el cuestionario, es necesario someterlo a prueba realizando una pe­queña **pre encuesta**, lo cual permitirá hacer las correcciones necesarias, mejorar su diseño tanto en las preguntas como en las instrucciones. La encuesta preliminar o piloto permite, entre otros aspectos, familiarizar al entrevistador con el formulario y con la unidad a la que se va aplicar el instrumento; conocer la reacción de los informantes; determinar el tiempo que dura la entrevista. De esta manera se podrá estimar el tiempo necesario que requiera la recolección. Hoy día, en cualquier investigación, es indispensable tener un mayor conocimiento sobre la población objeto de estudio, para poder estimar el grado de dispersión, para lo cual se aplica una de las medidas estadística denominada varianza o, en otros casos, la proporción, ambas fundamentales para determinar el tamaño ópti­mo de la muestra.

Una vez que el investigador ha estudiado los pros y contras en cada una de las pre­guntas, se organiza el cuestionario de forma apropiada y se alistará una prueba piloto, a fin de examinar el instrumento de recolección, en cuanto a claridad y extensión. Esa prueba se hace generalmente con un grupo pequeño de sujetos u objetos, y es una fase esencial para efectuar la encuesta. Este grupo de personas no solo aportarán un estima­tivo del tiempo necesario para responder a cada encuesta, sino que se les puede pedir sus comentarios respecto a ambigüedades percibidas y la recomendación de preguntas adicionales, si fuese necesario.

### Etapa 6. Recolección de datos

Distribuyendo o aplicando los instrumentos de recolección, de los cuales ya se hizo referencia, se recomienda la supervisión permanente del trabajo de los encuestadores, quienes con anterioridad recibieron un riguroso entrenamiento sobre la forma de en­trevistar, el manejo del instrumento y el propósito de la encuesta. Por otra parte, es el momento de volver a la fuente de información cuando se detecten inconsistencias, para así corregir fallas de algún entrevistador, o para hacer planes para las no respuestas o unidades que no fueron detectados o localizadas.

### Etapa 7. Organización y manejo de la información

Recolectada, verificada y controlada la información, debe precederse a su organización y manejo. La información podrá procesarse en forma manual o mecánica, dependiendo de la cantidad de unidades o de preguntas determinadas, de la disponibilidad de equipo o dinero. Esto permitirá agruparla de acuerdo con los criterios preestablecidos, acordes con la forma de presentación, análisis y descripción del hecho, de manera que faciliten la obtención de conclusiones o recomendaciones.

### Etapa 8. Análisis de los datos

Luego de organizada la información, se procederá al análisis de los datos comenzando por la aplicación de aquellas medidas que permitan la obtención de estimadores, y que además faciliten la comprobación de las hipótesis establecidas.

### Etapa 9. Redacción y entrega del informe final

Corresponde a la última etapa de la investigación y a la culminación de todos los trabajos que la misma causó.

A pesar de que el informe constituye un todo indivisible, se consideran tres partes perfectamente identificables: introducción, conclusiones y apéndices.

* **Introducción** En esta parte del informe se utilizará un lenguaje sencillo y ameno, dejando por fuera los tecnicismos, de manera que sea entendible por quien lo lea; además, debe mantenerse una secuencia que guarde armonía con el proceso utilizado en la investigación. Se hará claridad sobre el planteamiento del pro­blema que condujo a la realización del estudio, la fijación de objetivos, elección del método de investigación aplicado, incluido el método de selección utilizando; el tamaño de la muestra calculada, así como el diseño del cuestionario, las preguntas que se hicieron; la indicación de la fecha en que se inició y terminó la recolección de la información. En los capítulos siguientes, en forma breve y descriptiva, se hará la exposición de los puntos más importantes, apoyados en cuadros, gráficas y otros resultados que sean consecuencia de la aplicación de métodos estadísticos.
* **Conclusiones** Constituye la parte fundamental del informe, ya que en ella apare­cen relacionados los resultados obtenidos, la confrontación con lo esperado, la exposición de cómo deben ser interpretados y qué se deduce de ellos. El infor­me, además, debe remitir al lector, cuando sea necesario, a consultar todos los cuadros y gráficas que hagan más comprensible la explicación y, sobre todo, la presentación de recomendaciones.
* **Apéndices** Integra toda la documentación que se ha citado en la introducción y en las conclusiones, para que el lector del informe pueda consultar rápidamente y con facilidad cualquier información contenida en el mismo.

### Selección del tamaño de la muestra

En vez de realizarse un censo completo, los procedimientos del muestreo estadístico se han vuelto el instrumento preferido en la realización de la mayoría de las encuestas. Existen tres razones principales para preferir una muestra: primera, se suele necesitar demasiado tiempo para hacer un censo completo; segunda, es muy costoso; tercera, es engorroso e ineficiente obtener un conteo completo de la población objetivo. Por tanto, una encuesta ha de servir para hacer inferencia sobre la población total, con base en los resultados obtenidos de la encuesta, seleccionando una parte representativa de la población.

**Selección de elementos o unidades**

Cuando se hizo referencia al muestreo aleatorio simple (MAS), se mencionaron y explicaron, en forma simple, algunos de los métodos para la selección de unidades; sin embargo, vale la pena detenerse en ellos y dar una explicación un poco más amplia que permita su aplicación inmediata a través de ejercicios desarrollados en clases.

Para la utilización de cualquiera de los métodos de selección hay que tener en cuenta los aspectos que se describen a continuación.

* **Elaborar un listado completo y actualizado de los elementos o unidades que conforman la población objetivo** Por ejemplo: si se considera como población los alumnos matriculados en un plantel, si se va a investigar el total de droguerías existentes en la ciudad, a los trabajadores de una empresa o el total de manzanas de una zona o de un barrio, en todas estas poblaciones se deberá elaborar un listado o croquis que se constituye en el marco de referencia, y del cual se extraerán los elementos que conformarán la muestra. En la elaboración de unmarco, se podrá utilizar el libro o actas de matrícula que tenga el plantel; el listado de droguerías inscritas en la Secretaria de Salud o en la Cámara de Comercio; la nómina que utilizó la empresa para el pago de los últimos salarios. La Oficina de Catastro los puede suministrar el croquis donde se localizan las zonas, barrios o viviendas objeto de estudio.
* **Las unidades o elementos se identificarán en el marco, ya sea mediante el nombre, su dirección o el uso de un código** Generalmente, se procede a la enumeración de la población. Si esta está constituida por 82 elementos, se numerarán desde 01 hasta 82; si la conforman 480 elementos, se numerarán desde 001 hasta 480 (utilizando tres dígitos para cada cifra); si la población es de 1.350 elementos, se utilizarán cuatro dígitos, y así sucesivamente.
* **Seleccionar y aplicar el método más conveniente** Se hará la selección aplicando cualquiera de los siguientes métodos: sorteo, tablas de números aleatorios, selección sistemática, utilización de calculadoras o Excel.

1. **Sorteo** Se procede en una forma similar a la realización de un sorteo o al juego de lotería. Para ello, se preparan fichas, cartoncitos o balotas. Cada balota, por ejemplo, identificará un elemento. Introducidas en un recipiente, se extraen de una en una, con reposición o sin esta, revolviéndolas en cada extracción.
2. **Selección sistemática** Denominada también selección a intervalos regulares, se emplea de preferencia cuando las unidades están ordenadas en relación con el valor o el tiempo. En las oficinas, algunas veces, se cuenta con gavetas que contienen archivos diseñados en tarjetas, que permiten extraer varias de ellas a intervalos regulares. El procedimiento que se sigue es calcular un intervalo de selección:

*I*= N/n. Siendo *N*: el tamaño de la población; *n*: el tamaño de la muestra.

Supóngase que *N* = 80 y *n* = 10, el valor de *I* será 80/10 igual a 8. Luego, se selec­ciona un número al azar dentro de este intervalo, es decir, entre 1 y 5. Supóngase que se obtuvo el 3. Este será el primer elemento seleccionado; el segundo será 11; el tercero 19, y así sucesivamente irá sumándosele a cada número resultante el valor del intervalo, hasta completar el valor correspondiente al tamaño de la muestra.

1. **Uso de las tablas de números aleatorios o tablas Random** Este método aún uti­lizado, garantiza que todos los elementos tengan igual probabilidad de ser seleccio­nados, pues se eliminan aquellas decisiones negativas como el factor humano y las condiciones físicas del material utilizado en la elaboración de las fichas. Estas tablas se presentan de forma diferente, con cifras de 2, 4, 6, 8 o 10 dígitos, con el fin de hacer más fácil su observación, como se muestra en la tabla 1.1. Para la lectura de esta tabla, se recomienda:

* Dentro de un conjunto de tablas al azar seleccionar una de ellas (en este caso la tabla 1.1).
* Luego, sin mirar la tabla seleccionada, poner el dedo en un lugar cualquiera, que se considera como punto de partida.
* Determinar cuántos dígitos se tendrán que utilizar, de acuerdo con el tamaño de la población.
* Establecer la dirección que se seguirá en su lectura.
* Solo se seleccionarán aquellos números que formen parte del marco muestral.
* Hacer la selección con reposición o sin esta. En el primer caso, el elemento podrá seleccionarse más de una vez; en el segundo, una vez tenido en cuenta, no podrá seleccionarse nuevamente.

**Ejemplo:**

Considérese la selección de los elementos o unidades en una muestra aleatoria mediante el uso, en primer lugar, de la tabla de números aleatorios o tabla Random. Para ello, se supone disponer de una población hipotética de 250 elementos, de la cual se hace necesario contar con una muestra aleatoria de 25, para el estudio sobre duración, resistencia, precio, entre otras variables requeridas.

**Solución**

En primer lugar se le asigna un número a cada elemento, en este caso, desde 001 hasta 250. Luego se selecciona una tabla que contenga números al azar.

De esta forma se determina el punto de partida; por ejemplo, la fila 16 y la columna 11, y se obtendrá el primer número: 524. Si se lee de izquierda a derecha, el segundo será 962; luego, 094, 080, y así sucesivamente. También se hubiese podido obtener cualquiera de los siguientes números: 524, 249, 496, 620, 209, al realizarlo en forma continua formando cifras de tres dígitos, Si la lectura se hiciera por columnas se tendrían: 524, 601, 100, 557, siendo 421 el último en la columna. La forma de continuar depende de lo establecido antes de iniciar la selección, es decir, si se toma en la parte inferior, el número será 681; luego el 044, y así sucesivamente hacia arriba y en esa misma columna; o por el contrario, haber comenzado desde la parte superior con los números 646, 676, 686...

# TEMA 2: SUMATORIA

La sumatoria corresponde a una de las tantas operaciones que se realizan en la estadís­tica, ya que siempre se tendrá que sumar numerosos términos, los cuales muchas veces deben expresarse por símbolos o fórmulas, con el fin de simplificar su presentación.

El signo de la sumatoria es  (sigma), letra griega que indica suma de o sumatoria de.

Por definición se tendrá que:



En términos generales, una operación de sumatoria se expresa así:



Al signo sigma  la acompaña i = 1, que representa el límite inferior, donde i no necesariamente debe ser igual a uno; n corresponde al límite superior y, finalmente, i es el elemento genérico de la suma. Una calculadora de bolsillo tiene algunas teclas que al oprimirse, dan resultados para las expresiones:

* 
* 
* 
* 
*  y n, que corresponde al número de observaciones o datos registrados.

La expresión  se lee: "Sumatoria de i cuando i va de 1 hasta n ", lo cual quiere decir que i toma valores desde uno hasta n y el resultado final será la suma de esos valo­res. Algunas aplicaciones:

1.  = 1 + 2 + 3 + 4 + 5 = 15
2.  = 1+ 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55

En un ejercicio igual a los anteriores, si n es demasiado grande se podrá simplificar mediante la aplicación de una de sus propiedades.

## Propiedades de la sumatoria

1. 

* 

1. 

* 

Pues si no se aplica esta propiedad, el proceso es más dispendioso, veamos:



1. 

* 

Como se observó en la aplicación de la propiedad anterior, en esta se observa con mayor claridad la forma como se simplifican las operaciones.



1. En el caso de la sumatoria de una constante, es más importante la aplicación de las propiedades, por ejemplo:

= 10 + 10 + 10 + 10 + 10 = 50

Siendo n = 5 y k = 10, se podrá realizar la operación mediante el producto de nk = 5(10) = 50, y la propiedad se expondrá así: "La sumatoria de una constante es igual al producto del límite superior n por la constante K".

 = n \* k

El procedimiento anterior se aplica cuando el límite inferior de la sumatoria es igual a uno; en caso contrario, se realiza una pequeña modificación, tal como se verá en el siguiente ejercicio:

= 12 +12 +12 +12 +12 = 60 entonces ((8 -4) +1)12 = 5(12) = 60

Como se observa en el anterior ejercicio, al límite superior se le resta el valor del imite inferior; a este resultado se le suma uno y el valor obtenido se multiplica por la constante:



1. Ahora, considérese la sumatoria de una constante por una variable.

 = 8(1) + 8(2) + 8(3) + 8(4) + 8(5) = 8 + 16 + 24 + 32 + 40 = 120

Aquí es más fácil aplicar la propiedad que dice: "La sumatoria de una constante por una variable es igual al producto de la constante por la sumatoria de la variable".



*  = 8(1 +2 + 3 + 4 +5) = 8(15) = 120

1. La sumatoria de dos o más términos, todos ellos incluidos dentro del paréntesis. Esto dará origen a tantas sumatorias dependiendo del número de variables que se encuentren dentro de él.



Ejemplo:

Resolver la siguiente sumatoria.





= 55 + 45 + 40

Solución:

Si la sumatoria cobija varios términos, pero ellos no se encuentran dentro del parén­tesis, el proceso será diferente al anterior.

Como se observa en el ejemplo anterior, la sumatoria solo cobijará al primer térmi­no y el desarrollo de esta expresión quedaría así: 55 + 3i + 8 o 63 + 3i

### Practica lo aprendido

1. Un ejemplo de característica cualitativa puede ser datos sobre:
2. Salarios.
3. Ocupación.
4. Gastos en alimentación.
5. Ausencias a clase.
6. Número de accidentes en el trabajo.
7. Cualquier medida aplicada a las características de las unidades o elementos en la población se denomina:
8. Estimador.
9. Variable.
10. Estadística.
11. Parámetro.
12. Ninguna de los anteriores.
13. La estadística descriptiva tiene como objetivo:
14. Probar la significación de los resultados.
15. Lograr conclusiones más allá de las muestras.
16. Ser aplicada en investigaciones poblacionales.
17. Ninguna de las anteriores.
18. Por población o universo se entiende:
19. Conjunto de seres humanos.
20. Un recuento de unidades.
21. Conjunto de medidas o recuento de todas las unidades que tienen una característica común.
22. Es un ordenamiento sistemático de la información.
23. Descubrir las causas que originan el hecho.
24. Determine cuáles de las siguientes variables son cualitativas o cuantitativas y entre estas últimas cuáles son discretas o continuas:
25. Vida útil de una bombilla.
26. Marca de una bombilla.
27. Cotización de unas acciones.
28. Número de accidentes en una fábrica.
29. Tipos de accidentes que ocurren en una fábrica.
30. Estado civil de los trabajadores.
31. Lea el siguiente caso:

En un barrio de clase media de una capital, se localizó un total de 320 manzanas: además, cada manzana (aproximadamente) cuenta con 30 viviendas. El objetivo de la investigación es determinar el gasto promedio mensual en artículos de aseo. Se estableció la necesidad de realizar una selección de 40 manzanas y en cada una de ellas seleccionar 10 viviendas.

Algunas preguntas que contiene el formulario:

* Cantidad de dinero gastado.
* Establecimientos donde generalmente compra.
* Día que generalmente lo hace.
* Tiempo (horas) dedicado a sus compras.
* Número de veces que realiza el mayor número de compras.
* El número de personas que conforman su hogar.

Determine los siguientes aspectos:

1. Método de muestreo utilizado.
2. Población.
3. Muestra (tamaño).
4. Marco.
5. Elemento o unidad.
6. En las características investigadas, cuáles son atributos y cuáles son variables; en estas últimas identifique las variables discretas y continuas.
7. En cada uno de los siguientes problemas identifique: la población, el elemento o unidad. Establezca como mínimo dos características que podrían estudiarse.
8. Un fabricante de helados tiene varios puntos de venta en la ciudad de Cúcuta y desea estudiar si existe una relación o no, entre el valor de las ventas y la temperatura ambiental.
9. El director de recursos humanos de una universidad desea estudiar si existe o no relación entre la ausencia al trabajo (número de días) durante el año y la edad del empleado.
10. Una promotora de vivienda social está interesada en establecer cómo el aumento en los materiales de construcción está afectando la construcción de vivienda en una región del país. Como parte de la investigación planea reunir datos realizando una muestra de empresas constructoras durante los últimos cinco (5) años.
11. Una empresa distribuidora de gasolina desea determinar los factores que influyen en la elección de las estaciones de gasolina (Shell, Texaco, Esso, Terpel, Petrobrás), por los conductores de vehículos en la ciudad, motivo por el cual desea hacer una investigación por muestreo.
12. ¿Cómo podría hacerla?
13. ¿Qué preguntas podría formular?
14. ¿Qué unidad utilizaría?
15. ¿Qué método de muestreo sería el indicado?
16. Un auditor se enfrenta en una larga lista de cuentas por cobrar en su empresa. Debe verificar las cantidades (valor) con base en una muestra que abarca aproxi­madamente el 10% de esas cuentas, con el fin de estimar la diferencia promedio entre los valores revisados (muestra) y sus respectivos asientos en el libro.
17. Suponga que las cuentas están ordenadas cronológicamente, admitiendo que las más antiguas tienen la tendencia a tomar valores pequeños. ¿Qué método de muestreo utilizaría?
18. Suponga que las cuentas no están ordenadas. ¿Cuál sería el método de muestreo en este caso?
19. Suponga que las cuentas están divididas en tres (3) grupos de acuerdo con su valor. ¿Cuál sería el método por utilizar?
20. Se tiene un grupo de consumidores e interesa conocer sus preferencias por uno de estos productos que la empresa fábrica, que son de consumo popular. De acuerdo con el objetivo, se ha diseñado una muestra, para lo cual se selecciona un barrio, conformado aproximadamente por 120 manzanas, cada una de ellas con un prome­dio de 28 viviendas.
21. Formule el procedimiento que se utilizó en la selección. En primer lugar 18 manzanas y luego de cada una de ellas, 4 viviendas. ¿Qué método se empleó?
22. Si el método que le han exigido que realice es de un muestreo aleatorio simple, ¿cómo procedería?
23. De las siguientes características, cuáles considera: atributo o cualitativa (A); cuantitativa (C); variable discreta (VD); variable continua (VC).
24. Número de veces que los alumnos de una universidad utilizan el servicio de transmilenio, en un mes.
25. Alumnos atendidos por el servicio médico que ofrece la universidad en el mes, según diagnóstico.
26. Gasto semanal en que incurre un estudiante.
27. Número de veces que en el mes frecuenta la biblioteca.
28. Deportes que más practican los estudiantes los fines de semana.
29. Explique brevemente qué entiende por:
30. Dominio de estudio y de qué manera puede ocurrir.
31. Marco muestra y marco defectuoso.
32. Sustitución de elementos o unidades. ¿De cuántas maneras se pueden sustituir cuando no hay información?
33. Preguntas cerradas, abiertas, de filtro, de control, de selección múltiple.
34. Hechos que abarca y no abarca la estadística.
35. Explique brevemente qué entiende por:
36. Estadística descriptiva e inferencia estadística.
37. Estadísticas internas y externas.
38. Estadísticas primarias y secundarias.
39. Error de muestreo y ajeno al muestreo.
40. Investigación total y parcial.
41. Estadística y estadísticas.
42. Explique brevemente:
43. ¿Qué factores impiden realizar un censo o investigación total?
44. Los métodos de muestreo aleatorio y no aleatorio.
45. Las ventajas del muestreo.
46. Finalidad de la estadística.
47. ¿Cuándo un método de muestreo no aleatorio puede considerarse válido?
48. En cada uno de los casos señalados, en la siguiente lista, indique si es variable o atributo; al mismo tiempo, establezca la medida más indicada (media total, proporción).
49. Biología: peso de un ser vivo.
50. Agricultura: cultivo de legumbres.
51. Economía: ventas efectuadas por una empresa.
52. Transporte: en ferrocarril (en toneladas-km).
53. Demografía: número de hijos de una familia.
54. Demografía: estado civil.
55. El directorio de una ciudad data de 4 años atrás y tiene las direcciones a lo largo  
    de cada calle, con los nombres de las personas que viven en cada una de ellas.
56. ¿Cuáles son las deficiencias de este marco si se utiliza para tomar una encuesta por muestreo?
57. ¿Pueden corregirse tales deficiencias durante los trabajos de terreno?
58. Al usar el directorio, ¿extraería usted una lista de direcciones o una lista de personas?
59. Explique brevemente qué diferencias puede haber entre:
60. Estadística y estadísticas.
61. Parámetro y estimador.
62. Variable discreta y continua.
63. Característica cuantitativa y cualitativa.
64. Mencione algunos aspectos relacionados con:
65. Las limitaciones que hay para la realización de un censo.
66. La finalidad en la aplicación de la estadística.
67. Dé tres ejemplos de medidas aplicadas en la estimación de las características de las unidades de una población: media aritmética, proporción, total y razón.
68. Utilizando como población los estudiantes de su plantel, identifique 10 características. Diga cuáles son atributos y cuáles variables. En las variables distinga las discretas y las continuas.
69. Señale con una (A) las series constituidas por datos cualitativos y con una (V) las que son cuantitativas.
70. Nacionalidad.
71. Temperatura.
72. Estado civil.
73. Producción.
74. Número de tornillos producidos.
75. Filiación política.
76. Kilómetros recorridos.
77. Gasto en alimentación.
78. Llamadas diarias.
79. Cuáles son las limitaciones y en qué circunstancia es preferible la realización de un censor
80. Los directivos de una fábrica estudian un programa de vivienda para sus empleados y solicitan información a un grupo de 30 de ellos, de un total de 350 emplea­dos, sobre el número de habitaciones que requerirá su vivienda, en el caso de salir crecido en el programa. La investigación se hace únicamente en aquellos que tienen más de cuatro años de antigüedad y en la actualidad no tienen vivienda. Determine:
81. La población.
82. La muestra.
83. La característica y cómo la clasificaría.
84. Cuáles de los siguientes casos corresponden al muestreo aleatorio:
85. Si de las jaulas de un criadero de gallinas se toma una muestra de 25 animales, seleccionando aquellos que alcanza la mano.
86. Si se toman las hojas de registro de matrícula de los alumnos de un instituto y se selecciona una muestra de 20 hojas a intervalos regulares.
87. Si de una muestra de frutas que se encuentran en un recipiente bastante grande, se toma como muestra la porción correspondiente a los 10 centímetros de la parte superior de la vasija.
88. Si para conocer el sentimiento popular sobre una medida de orden público, se entrevistan a 30 personas que pasan por un sitio importante de la ciudad.
89. Escriba en forma explícita las sumas representadas por cada una de las siguientes expresiones.
90. 
91. 
92. 
93. 
94. 
95. 
96. Escriba cada una de las siguientes expresiones, utilizando el signo de la sumatoria. con los límites adecuados.
97. X1 + X2 + X3 + X4
98. 
99. 
100. 
101. Si X1 = 3, X2 = 9, X3 = - 7, X4 = -3, calcule el valor numérico de las siguientes expresiones.
102. 
103. 
104. Si X1 = 8, X2 = 4, X3 = 4, X4 = 0, calcule:
105. 
106. donde a = 4.
107. En los ejercicios siguientes, suponga que se da un conjunto de números: X1, X2, X3,…, Xn y que:



Demuestre las siguientes relaciones.

Nota: cuando se da la sumatoria sin los límites inferior y superior, se entiende que el límite inferior es igual a = 1 y el límite superior será n, tal como ocurre con los ejerci­cios (b) y (d).

1. Resuelva
2. 
3. 
4. 
5. 
6. 
7. 
8. 
9. 
10. 
11. 
12. 
13. 
14. 
15. 
16. 
17. 
18. 

Nota: considere los siguientes valores de Xi y Yi para los ejercicios d, e, f, o y  
p:

* X1 = 8
* X2 = 0
* X3 = 5
* X4 = 2
* X5 = 3
* X6 = 4
* Y1 = 2
* Y2 = 3
* Y3 = 6
* Y4 = 2
* Y5 = 7
* Y6 = 5

# TEMA 3: DISTRIBUCIONES DE FRECUENCIAS

## PRESENTACIÓN DE DATOS

Una vez recolectada toda la información, por cualquiera de los métodos (censo, muestra o registro continuo), tal como sucede en las diferentes investigaciones, se debe proceded a su organización, mediante dos procesos:

1. Clasificación u ordenamiento de los datos.
2. Tabulación o procesamiento de la información.

## CLASIFICACIÓN U ORDENAMIENTO DE LOS DATOS

Consiste en determinar el criterio que se va a utilizar, para agrupar los datos recolecta­dos. Algunos de los criterios que pueden tenerse en cuenta son:

* Orden alfabético: si se hace una investigación en una empresa sobre su personal, referente a cargos, profesión, sexo, departamento en el que labora, etcétera, se observará a primera vista que todas estas características son cualitativas y se pue­den clasificar alfabéticamente.
* Geográfico: si al personal de esa empresa se agrupa de acuerdo con el lugar de procedencia o de nacimiento (no necesariamente en orden alfabético), la información presentará por regiones o zonas (por ejemplo, de norte a sur, de oriente a occidente),
* Histórico o cronológico: se denomina así, si la característica se representa por días, meses o años; por ejemplo, los accidentes de tránsito, según las causas que originan, se pueden ordenar alfabéticamente, geográficamente o por periodos de tiempo, es decir, en forma cronológica.
* Convencional: muy utilizado por el Departamento Administrativo Nacional de Estadística (DANE) obedece a recomendaciones internacionales, a fin de que la información guarde uniformidad con la de los demás países.
* Usual: cuando se emplean sin ningún criterio especial, solo por costumbre.
* Numérico: por ejemplo, el número de alumnos ordenados por grados 1, 2... 11.

### Formas de medición

Se vio que las características se dividen en cualitativas y cuantitativas. Las cualitativas se clasifican utilizando escalas nominales y las cuantitativas, ordinales y numéricas. Escalas ordinales Supóngase la realización de una investigación en torno a fumadores sobre las marcas preferidas; los resultados fueron: Andino, Derby, Kool, Lucky Strike. Marlboro y Pielroja. La característica "Marca de cigarrillos" queda clasificada por las sie­te categorías encontradas. Si con la anterior información se clasifica a los fumadores por nivel socioeconómico dentro de tres categorías (alto, medio y bajo), se estará utilizando una escala ordinal, puesto que se ha tenido en cuenta su orden jerárquico.

1. Escalas numéricas Son aquellas que se utilizan para clasificar las características cuan­titativas o variables. Algunos las dividen en escalas a intervalos regulares y escalas pro­porcionales.

* Escalas a intervalos regulares: se establecen cuando no existen patrones de medi­ción de uso común, por ejemplo, cuando se requiere medir la inteligencia, los hábitos, actitudes, formas de comportamiento.
* Escalas proporcionales: al contrario de las anteriores, están dadas por unidades de uso corriente, como cuando se clasifican 50 establecimientos de acuerdo con los volúmenes de ventas, número de empleados, etcétera.

1. Tabulación u organización de los datos

Después de determinar el criterio de clasificación que se va a seguir, hay que iniciar el proceso de tabulación, el cual consiste en un recuento de las veces que se repite la característica observada: por ejemplo, ¿cuántos de los Fumadores prefieren cigarrillos de la marca Andino?, ¿cuántos de la marca Derby?, y así sucesivamente.

La tabulación puede realizarse en forma manual o mecánica. La decisión depende de 35 factores, entre otros: del número de formularios, de la cantidad de preguntas que tenga el instrumento de recolección y de la disponibilidad de equipo para procesar la información. De todas formas, se requiere elaborar un plan de tabulación que debe referirse al diseño de una serie de cuadros o tablas, los cuales deben atender al objeto de la encuesta y permitir así la presentación de los datos en forma clara, que facilite su análisis y la aplicación de algunas medidas de carácter estadístico, como se verá más adelante, y en la mayoría de los casos acompañados de una representación gráfica.

### Cuadros

Los cuadros que constituyen el plan de procesamiento o sistematización, sea manual o mecánica, no deben ser necesariamente iguales a aquellos que finalmente se van a presentar o publicar; sin embargo, es preferible que sean iguales por economía de tiempo y dinero. Los cuadros, de acuerdo con el número de características que se van a presentar, se clasifican en simples, de doble y múltiple entrada.

* Cuadros simples: siempre representan una sola variable o un atributo.
* Cuadros de doble y múltiple entrada: se elaboran para presentar dos o más carac­terísticas. Cuando en un cuadro se combinan variables con atributos, algunos lo denominan cuadro o tabla mixta.

En la elaboración de cuadros finales o de publicación, se recomienda:

1. Si el informe tiene más de un cuadro, los cuadros deben numerarse de forma consecutiva.
2. Todo cuadro debe tener un título, lo más completo posible, de manera que al leerlo se tenga una visión general sobre su contenido.
3. La mayoría de los técnicos recomiendan que el título debe situarse en la parte supe­rior del cuadro; sin embargo, otros consideran que debe ir en la parte inferior.
4. Las líneas que encierren el cuadro deben ser más gruesas que las internas. Algu­nos aconsejan, por presentación, cerrar el cuadro por encima y por debajo, es decir, suprimir las líneas laterales.
5. Los títulos en las columnas deben ir en mayúscula y los subtítulos en minúsculas.
6. Toda observación que se haga debe indicarse en el sitio correspondiente con un llamado, y al final, debajo del cuadro, hacer la aclaración. En ese mismo sitio se acostumbra colocar la fuente de la cual se tomó la información.

## TABLAS DE FRECUENCIAS

Las tablas de frecuencias corresponden a una agrupación o clasificación de los datos con el fin primordial de determinar la frecuencia con que se repite el atributo o el valor que toma la variable. En este último caso, se refiere como distribución de frecuencias a aquellas tablas que se elaboran atendiendo al número de observaciones o datos relativamente grandes, siendo a la vez la aplicación primaria de la estadística.

**Atributos**

Se anotó que estas características se expresan también mediante palabras. El termine emplea habitualmente para distinguirlo de la palabra variable o característica cuantitativa. Por ejemplo, sexo, estado civil, profesión, cargo, ocupación, aficiones, estudios realizados y otros más, se clasifican como atributos; pero la edad, el salario, tiempo de servicio, personas a cargo, retardos en el mes, horas laboradas, etcétera, son ejemplos de variables.

Con frecuencia, los atributos son dicotómicos, de modo que cada miembro de una población se incluye en uno de los dos grupos, conforme posea una característica específica o no la posea. También permite clasificaciones múltiples, ya que una persona puede tener preferencia por más de una opción, por ejemplo, las marcas de un producto o deportes que práctica, etcétera. La tabulación es muy sencilla, pues su cuantificación se hace mediante el recuento del número de veces que se repite la característica investigada. Hoy día se ha generalizado el término de variable, para indicar tanto a las características cualitativas como a las cuantitativas, por tanto, no hay dificultad al usarlo, sin embargo se considera lógico la división anterior.

**Tabla 1:** Número de fumadores de acuerdo con la marca preferida

| Marcas | Fumadores | % |
| --- | --- | --- |
| Andino | 12 | 13,35 |
| Derby | 28 | 31,11 |
| Kool | 10 | 11,11 |
| Lucky | 7 | 7,78 |
| Marlboro | 20 | 22,22 |
| Pielroja | 13 | 14,45 |
| Total | 90 | 100 |

Lo anterior puede explicarse al realizar una encuesta a fumadores sobre su preferencia por una marca, cuyos resultados se muestran en las tablas siguientes. En estas se relaciona la información correspondiente a dos preguntas diferentes realizadas a 90 fumadores encuestados. La primera hace referencia a la marca de cigarrillo preferido, es decir, solo se debe señalar una sola marca, y el resultado total de fumadores es de 90 y la distribución porcentual de la suma debe ser el 100%; en cambio, en la segunda pregunta, se hace referencia a la marca de cigarrillos que generalmente fuma, y la suma debe ser superior a los 90 encuestados y, por consiguiente, el porcentaje es mayor que el 100 %.

### Tablas o cuadros de frecuencias y representación gráfica

Ello se debe a que existen fumadores que tienen preferencias por más de una marca. Las dos presentaciones son totalmente distintas.

En ambas tablas, cada frecuencia se divide por 90. De esta forma se calculan las frecuencias; además, los datos de la primera tabla se podrán representar mediante una circular o de barras; la segunda solo podrá representarse mediante la gráfica de barras.

Una tabla de doble entrada podría ser la clasificación de 150 empleados, de acuerdo al estado civil y sexo.

Tabla 2: Número de fumadores de acuerdo con las marcas que compra generalmente

| Marcas | Fumadores | % |
| --- | --- | --- |
| Andino | 16 | 17,78 |
| Derby | 36 | 40,00 |
| Kool | 12 | 13,33 |
| Lucky | 16 | 17,78 |
| Marlboro | 28 | 31,11 |
| Pielroja | 22 | 24,44 |

Tabla 3: Clasificación de empleados de una empresa, según estado civil y sexo

| Estado Civil | Masculino | Femenino | Total |
| --- | --- | --- | --- |
| Soltero | 28 | 22 | 50 |
| Casado | 52 | 10 | 62 |
| Separado | 17 | 5 | 22 |
| Viudo | 3 | 1 | 4 |
| Unión libre | 10 | 2 | 12 |
| Total | 110 | 40 | 150 |

## VARIABLES

Estas se clasifican en discretas y continuas. Las primeras son todas aquellas que admiten lente valores enteros; las segundas aceptan valores fraccionarios. El procedimiento que se debe seguir en la elaboración de una tabla de frecuencias se hará atendiendo a la anterior clasificación.

### Variable discreta

En la elaboración de una tabla de frecuencias, para variables discretas y continuas, se utiliza una serie de símbolos con los cuales hay que familiarizarse, ya que continuamente emplearán, y se constituyen en una forma de lenguaje. Estos términos son:

* : Es una manera de simbolizar el valor de cada observación que toma la variable. Donde i puede tornar valores desde 1, 2,3... hasta n, cuando se trata de una muestra, o hasta N, si se refiere a una población, y corresponde a datos originales o no agrupados. En datos agrupados, cuando se ha elaborado una tabla de frecuencias, i toma valores desde uno hasta m.

Algunos utilizan la X mayúscula para observaciones en poblaciones y minúscula en la muestra.

* m: Es el número de valores que toma la variable en una tabla de frecuencias. En la variable continua recibe el nombre de número de marcas de clase o número de intervalos.

Las marcas de clase, en datos agrupados, es igual al promedio que se obtiene en cr­eí límite inferior y el superior de cada intervalo.

* n: Corresponde al tamaño de la muestra.
* N: Es el número de elementos que tiene una población.
* fi o ni Se denomina frecuencia absoluta y corresponde al número de veces que se repite cada valor de la variable, o dentro de los intervalos de los límites determinados. Siempre son números enteros, y la suma deberá ser igual al total de las observaciones investigadas, es decir, n en la muestra o N en la población.
*  o hi: Es Infrecuencia relativa, entendida como los cocientes entre el valor de una determinada frecuencia absoluta y el total de observaciones. Las frecuencias relativas serán valores fraccionarios, es decir, valores comprendidos entre 0 y 1.

Además, la suma de todas estas frecuencias debe ser igual a 1, es decir, el 100%.

* Fi o Ni: Frecuencia absoluta acumulada, la cual se determina si se desea el total de datos u observaciones, desde el origen hasta el valor total. Ello se obtiene acumulando o agregando frecuencias absolutas.
* Hi: Frecuencia relativa acumulada. Se obtiene en forma análoga a las frecuencias absolutas acumuladas.

Para entender mejor el uso de estos símbolos, un ejemplo de variable discreta permitirá construir una tabla de frecuencias.

**Ejemplo:**

Una empresa tiene 220 empleados, de los cuales se selecciona una muestra al azar de 50 de ellos, con el fin de investigar qué porcentaje y con qué frecuencia hay retardos en la llegada del personal en las horas de la mañana. Para ello, se recoge las tarjen de control de la última quincena, cuyos resultados se muestran en la siguiente tabla:

Tabla 4: Datos originales o no agrupados

| Dato | Valor | Dato | Valor | Dato | Valor | Dato | Valor | Dato | Valor |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 11 | 6 | 21 | 4 | 31 | 3 | 41 | 0 |
| 2 | 1 | 12 | 3 | 22 | 0 | 32 | 5 | 42 | 2 |
| 3 | 2 | 13 | 3 | 23 | 3 | 33 | 3 | 43 | 3 |
| 4 | 4 | 14 | 4 | 24 | 1 | 34 | 1 | 44 | 6 |
| 5 | 0 | 15 | 4 | 25 | 3 | 35 | 2 | 45 | 3 |
| 6 | 2 | 16 | 2 | 26 | 4 | 36 | 4 | 46 | 5 |
| 7 | 2 | 17 | 2 | 27 | 2 | 37 | 5 | 47 | 5 |
| 8 | 1 | 18 | 3 | 28 | 3 | 38 | 1 | 48 | 3 |
| 9 | 3 | 19 | 4 | 29 | 1 | 39 | 1 | 49 | 2 |
| 10 | 5 | 20 | 3 | 30 | 3 | 40 | 3 | 50 | 4 |

Cada elemento se representa en un símbolo. Así, si se pide cuáles son los valores para x11, x26 y x42, se dirá que son 6, 4 y 2, respectivamente.

Ahora, se determinan los valores que toma la variable y se observa que son:

0, 1, 2, 3, 4, 5, 6

En total 7 el cual se simboliza como m = 7. Vale aclarar que la variable puede presentarse de manera discontinua; por ejemplo, xi: 1, 3, 4, 6, 9 ,10; en este caso m = 6.

Luego se tabula, es decir, se determina cuántas veces se presenta o se repite, cada valor que toma la variable, como se muestra en la tabla siguiente.

Tabla 5: Tabulación. Variable discreta

| xi | Tabulación | Frecuencia |
| --- | --- | --- |
| 0 | 1 1 1 | 3 |
| 1 | 1 1 1 1 1 1 1 | 7 |
| 2 | 1 1 1 1 1 1 1 1 1 1 | 10 |
| 3 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 15 |
| 4 | 1 1 1 1 1 1 1 1 | 8 |
| 5 | 1 1 1 1 1 | 5 |
| 6 | 1 1 | 2 |
| Suma |  | 50 |

En las tablas siguientes se presentan, respectivamente, los valores de las frecuencias y sus símbolos.

Cuáles serían los valores en la tabla de frecuencias, para x3, f5, , N6 y H2.

En el mismo orden, los valores son:

* X3 = 2
* F5 = 8
*  = 0,14
* N6 = 48
* H2 = 0,20

Las frecuencias relativas se obtienen dividiendo las frecuencias absolutas por el nú­mero de observaciones; así; el resultado de la frecuencia relativa será:

1. 3  50 = 0,06
2. 7  50 = 0,14.
3. 10  50 = 0,20
4. 15  50 = .0,3
5. 8  50 = 0,16
6. 5  50 = 0,10, y así sucesivamente.

Tabla 6: Tabla de frecuencias. Variable discreta

| xi | fi | fi / n | Ni | Hi |
| --- | --- | --- | --- | --- |
| 0 | 3 | 0,06 | 3 | 0,06 |
| 1 | 7 | 0,14 | 10 | 0,20 |
| 2 | 10 | 0,20 | 20 | 0,40 |
| 3 | 15 | 0,30 | 35 | 0,70 |
| 4 | 8 | 0,16 | 43 | 0,86 |
| 5 | 5 | 0,10 | 48 | 0,96 |
| 6 | 2 | 0,04 | 50 | 1,00 |
| Suma | 50 | 1,00 |  |  |
| yi | ni | hi | Ni | Hi |

Tabla 7: Simbología. Tabla de frecuencias

| xi | fi | fi / n | Ni | Hi |
| --- | --- | --- | --- | --- |
| x1 | f1 | f1 / n | N1 | H1 |
| x2 | f2 | f2 / n | N2 | H2 |
| x3 | f3 | f3 / n | N3 | H3 |
| x4 | f4 | f4 / n | N4 | H4 |
| x5 | f5 | f5 / n | N5 | H5 |
| x6 | f6 | f6 / n | N6 | H6 |
| Suma | n | 1,00 |  |  |
| yi | ni | hi | Ni | Hi |

La frecuencia absoluta acumulada fue:

1. L primera: 3.
2. La segunda se obtuvo sumando la primera frecuencia más la segunda: 3 + 7 = 10
3. La tercera será igual a 3 + 7 + 10 = 20, o siendo igual a la suma de la segunda frecuencia absoluta acumulada con la tercera frecuencia absoluta, igual a 10, que era el acumulado, más 10, que es el valor de la frecuencia, igual a 20.

Algo similar se hace con la frecuencia relativa acumulada.

**Propiedades de las frecuencias**

* Las frecuencias absolutas son números enteros.
* La suma de las frecuencias absolutas será igual al número total de observador en este caso n.
* Las frecuencias relativas son números fraccionarios así: 0 < fi / n < 1.
* La suma de las frecuencias relativas será igual a 1.
* El último término de la frecuencia absoluta acumulada es igual al total de observaciones, en este caso n.
* El último término de la frecuencia relativa acumulada es igual a 1.

### Variable continua

Corresponde a aquella variable que admite fracciones, ya que son resultados de la aplicación de algún método de medición. A los símbolos anteriores dados para una varia discreta, es necesario agregarles algunos más, que son propios de esta variable.

* X máx.: Es el valor máximo que toma la variable, cuando los datos se encuentran sin agrupar.
* X mín.: Es el valor mínimo que toma esa variable.
* Rango = recorrido: Es la diferencia entre el valor máximo y mínimo que toma la variable.

R = X máx. – X mín.

* m: Número de intervalos o marcas de clase. Es un valor que, en algunos casos se toma por conveniencia o porque la información ha venido presentándose de esa misma forma. Se aconseja que el valor de m no debe ser menor que cinco ni mayor que 16, para otros hasta 18. Por uniformidad en los ejercicios de clases, es utilizar la fórmula:

m = 1 + 3,3 log n.

* C o i: Es la amplitud del intervalo y no necesariamente debe ser constante. Por facilidad, se utilizará una amplitud constante.

C = rango / m.

Siempre que su resultado sea fraccionario, por pequeña que ella sea, se aproximará al número inmediatamente superior.

**Ejemplo:**

Un estudio realizado a 50 establecimientos comerciales, sobre el volumen de ventas (millones $) en un mes cualquiera, se obtuvo el resultado que se muestra en la siguiente tabla:

Tabla 8: Datos no agrupados. Variable continua

| Dato | Valor | Dato | Valor | Dato | Valor | Dato | Valor | Dato | Valor |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 42 | 11 | 37 | 21 | 51 | 31 | 62 | 41 | 65 |
| 2 | 56 | 12 | 42 | 22 | 46 | 32 | 48 | 42 | 53 |
| 3 | 39 | 13 | 56 | 23 | 54 | 33 | 55 | 43 | 56 |
| 4 | 48 | 14 | 61 | 24 | 62 | 34 | 60 | 44 | 57 |
| 5 | 54 | 15 | 56 | 25 | 50 | 35 | 54 | 45 | 53 |
| 6 | 62 | 16 | 50 | 26 | 36 | 36 | 38 | 46 | 40 |
| 7 | 45 | 17 | 42 | 27 | 56 | 37 | 61 | 47 | 65 |
| 8 | 45 | 18 | 60 | 28 | 52 | 38 | 50 | 48 | 55 |
| 9 | 51 | 19 | 61 | 29 | 63 | 39 | 58 | 49 | 50 |
| 10 | 56 | 20 | 46 | 30 | 58 | 40 | 57 | 50 | 46 |

Los pasos que deben seguirse para la confección de una tabla de frecuencias, cuando una variable continua, son:

**Paso 1:**

Se determina el valor máximo y mínimo que toma la variable.

* X máx. = 65, o sea la observación x47 y X min= 36, correspondiente a la observación x26.

**Paso 2:**

Se obtiene el rango o recorrido.

* R = X máx. – X mín. = 65 - 36 = 29.

**Paso 3:**

Se determina el valor de m o número de intervalos necesarios para agrupar los datos.

Se puede elegir un valor arbitrario, si así se desea. Recuérdese que no deben ser muchos ni pocos los intervalos o marcas de clase; generalmente, debe utilizarse un número no mayor de 16, para otros 18, ni menor a 5.

También, obsérvese si la empresa o entidad que realiza el estudio utiliza estos procedimientos en las publicaciones o estudios anteriores. En estos casos, se deberá emplear tablas similares, con el fin de continuarlas presentando de la misma forma, con fines comparativos.

En este ejercicio se puede utilizar la fórmula m = 1 + 3,3 log n, para calcular el número de intervalos o marcas de clase.

Siendo m = 1+ 3,3 log 50 = 1 + 5 (1,698970) = 1 + 5,60660 = 7.

En este caso, como en los demás, siempre se aproximará al número inmediatamente superior, pero no tiene importancia si se trabaja con el valor inferior, en este caso 6. A pesar de que muchas personas, quizás por abreviar, calculan este valor mediante la aplicación de la fórmula m =  = 7,07 = 8, el procedimiento se considera como poco recomendable.

**Paso 4:**

Se determina el valor de la amplitud, es decir, la diferencia que hay en cada uno de los intervalos de clase y que se simboliza por C o i. Por comodidad y presentación debe ser un número entero; y en la mayoría de los casos, se le puede considera constante para cada uno de los intervalos de clase.

C =  =  = 4,14 = 5.

De todas formas, por pequeña que sea la fracción, siempre se debe aproximar al número inmediatamente superior. Si se mantienen constantes los valores de m = 7 y C = 5, se nota que el valor del rango cambia.

5 = rango  7, el nuevo rango será 35, es decir, se incrementó en 6 unidades.

Siempre que esto ocurra, el valor excedente deberá repartirse, sumándole una parte al límite superior y restando la otra al límite inferior.

En este caso, como el incremento fue 6, se le suma 3 a 65 y se obtiene 68 como nuevo X máx., y si se le resta 3 a 36, lo cual dará 33 como X mín. El nuevo rango, en este caso, será igual a:

68 - 33 = 35.

Lo más indicado, en este proceso, es repartir el incremento hasta donde sea posible en partes iguales, pero no necesariamente se debe hacer.

**Paso 5:**

Una vez se establezca el nuevo rango, se coloca como Límite inferior del primer intervalo al valor mínimo del rango, en este caso 33, y va sumándosele el valor de la amplitud.

En la tabla siguiente, obsérvese que al límite inferior se le ha agregado 0,1, pero solo se hace con el fin de clasificar los datos. Es decir, si un valor de xi es igual a 48, quedar; dentro del tercer intervalo, ya que el cuarto corresponderá a valores superiores a 48.

Tabla 9: Tabulación. Variable continua

| Intervalos | Tabulación | Frecuencia |
| --- | --- | --- |
| 33,1 – 38 | 1 1 1 | 3 |
| 38,1 – 43 | 1 1 1 1 1 | 5 |
| 43,1 – 48 | 1 1 1 1 1 1 1 | 7 |
| 48,1 – 53 | 1 1 1 1 1 1 1 1 1 | 9 |
| 53,1 – 58 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 15 |
| 58,1 – 63 | 1 1 1 1 1 1 1 1 1 | 9 |
| 63,1 – 68 | 1 1 | 2 |
| Total |  | 50 |

Los símbolos que se utilizan, cuando se confecciona la tabla de frecuencias se muestran en la tabla siguiente, en la cual se calculan, a la vez, las marcas de clase.

Tabla 10: Simbología. Variable continua

| xi-1 - xi | fi | fi / n | Fi | Hi | xi |
| --- | --- | --- | --- | --- | --- |
| x0 - x1 | f1 | f1 / n | F1 | H1 | x1 |
| x1 - x2 | f2 | f2 / n | F2 | H2 | x2 |
| x2 - x3 | f3 | f3 / n | F3 | H3 | x3 |
| … | … | … | … | … | … |
| xm-1 - xm | fm | fm / n | Fm | Hm | xm |
| yi-1 - yi | ni | hi | Ni | Hi | yi |

Las marcas de clase que aparecen en la última columna de la tabla, se pueden calcular de varias maneras:

* Promediando los valores para cada intervalo. Para ello se le suma al límite inferior el valor correspondiente del límite superior del intervalo y luego se divide entre dos:

1.  = 35,5
2.  = 40,5, y así sucesivamente.

* Si la amplitud es constante, basta obtener la primera marca de clase por el método anterior y va sumándosele el valor de la constante, que en este caso es igual a 5.
* Se divide la constante por 2 y el resultado se le suma al límite inferior, o se le resta al límite superior del intervalo.

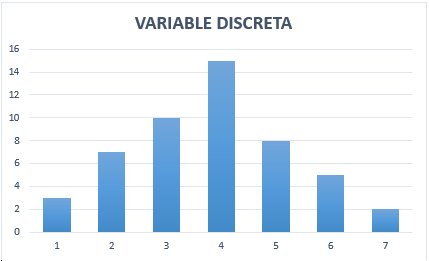
# TEMA 4: GRÁFICAS CON APLICACIONES EN LA TEORÍA ESTADÍSTICA

En el desarrollo de la teoría estadística, se emplean algunas gráficas que ayudan a visualizar y aplicar mejor el contenido de las tablas; estas son: diagramas de frecuencias, polígonos y ojivas.

## DIAGRAMAS DE FRECUENCIAS

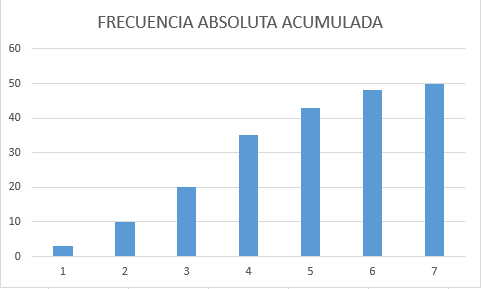
Por lo general se utilizan para representar las variables discretas, por medio de líneas ver-n altura estará dada por los valores de las frecuencias, ya sea absoluta o relativa; de líneas horizontales que dan la sensación de peldaños en una escalera.

Gráfica 1: Diagrama de frecuencia Xi



**Descripción Imagen:** Gráfico en el que se muestran barras con altura dado por la frecuencias de la variable discreta, la primera barra tiene altura 3, la segunda 7, la tercera 10, la cuarta 15, la quinta 8, la sexta 5 y la séptima 2.

**Gráfica 2:** Diagrama de frecuencias absolutas acumuladas (variable discreta)



**Descripción Imagen:** Gráfico en el que se muestran barras con altura creciente dado por la frecuencias acumuladas de la variable discreta, la primera barra tiene altura 3, la segunda 10, la tercera 20, la cuarta 35, la quinta 43, la sexta 48 y la séptima 50.

Como puede observarse en las gráficas anteriores, utilizando un par de ejes cartesia­nos, en el eje horizontal o abscisa se señalan los distintos valores correspondientes a la variable discreta, y en el eje vertical u ordenada, las frecuencias absolutas, relativas o la acumulada. Cada valor de la variable, con su correspondiente frecuencia, constituí una pareja que se señalará en el plano mediante un punto, desde los cuales se tra­zan líneas verticales u horizontales que, en conjunto, conforman gráficas denominadas diagramas de frecuencias.

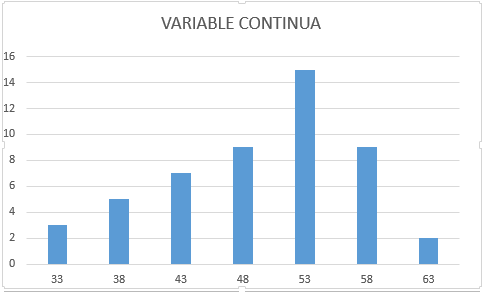
## HISTOGRAMAS, POLÍGONOS Y OJIVAS

Estas gráficas se utilizan para representar una variable continua.

### Histograma de frecuencias

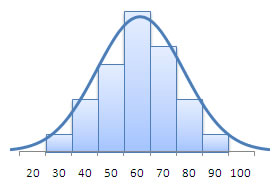
Para su elaboración requiere que la amplitud del intervalo sea constante; se compone de barras o rectángulos unidos, levantados sobre su base o abscisa. Su ancho está dado por el intervalo de clase y su altura por la frecuencia, absoluta o relativa.

**Gráfica 3:** Histograma de frecuencias



**Descripción Imagen:** Gráfico en el que se muestran barras con altura dada por las frecuencias de la variable continua, la primera barra tiene altura 3, la segunda 5, la tercera 7, la cuarta 9, la quinta 15, la sexta 9 y la séptima 2. Esta gráfica es un histograma, usado para variable continua.

Gráfica 4: Curva normal



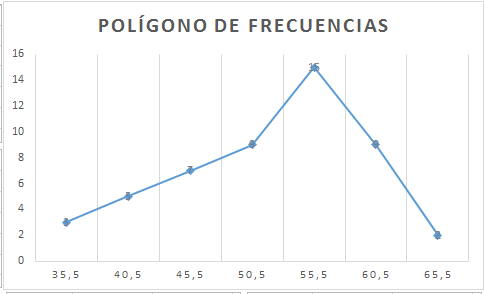
**Descripción Imagen:** Gráfico en el que se muestran barras con altura dada por las frecuencias de la variable continua, sobre las barras, en los extremos superiores se encuentra la graficada una curva suave en forma de campana.

### Polígono

Puede dibujarse sobre el histograma o aparte, si así se desea. En el primer caso, se unen superiores en cada barra o rectángulo; en el segundo, se unen los puntos de intersección de la abscisa, correspondiente a la marca de clase, y la ordenada (frecuencias relativas).

Cuando la distribución es simétrica, el polígono toma la forma de una campana, denominada generalmente de Gauss, ya que los valores de las frecuencias absolutas y relativas equidistantes a un valor central son iguales, creciendo desde los valores extremos hacia el centro. Esta gráfica es de gran importancia en el desarrollo de la teoría estadística y con grandes aplicaciones en la inferencia estadística.

Gráfica 5: Polígono de frecuencias Xi

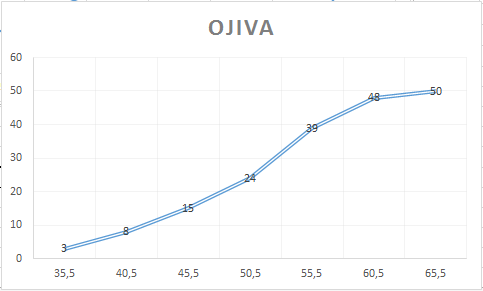


**Descripción Imagen:** Gráfico en el que se muestran la uniones mediante segmentos de recta de los puntos medios de las en su altura de las barras del histograma de frecuencias de la variable continua propuesta como ejemplo en el texto, este gráfico se denomina polígono de frecuencias.

### Ojiva

Se construye utilizando las frecuencias absolutas y relativas acumuladas, y se aplica para obtener soluciones en aquellos problemas donde hay necesidad de aplicar los percentiles.

Gráfica 6: Ojiva de Frecuencias.



**Descripción Imagen:** Gráfico en el que se muestran las uniones mediante segmentos de recta de los puntos medios, en la altura de las barras del histograma de frecuencias acumuladas de la variable continua propuesta como ejemplo en el texto, este gráfico se denomina ojiva.

Para la interpretación de estas gráficas, se considera únicamente el histograma de frecuencias. Considerando la gráfica de histograma se observa que la altura corresponde a la frecuencia absoluta o a la relativa; se dirá, entonces, que el 6% de los establecimientos, o sea 3, vendieron entre 33,1 y 38 millones de $; el 10% de los establecimientos, o sea 5-vendieron entre 38,1 y 43 millones de $; el 14% de los establecimientos, o sea 7, ven­dieron entre 43,1 y 48 millones de $; el 18% de los establecimientos, o sea 9, vendiera» i entre 48,1 y 53 millones de $.

## GRÁFICAS UTILIZADAS EN LOS INFORMES

Las gráficas presentadas anteriormente (diagramas de frecuencias absolutas, relativas y acumuladas), histogramas, ojivas y polígonos de frecuencias, como ya se anotó se aplican comúnmente en el desarrollo de la teoría estadística, pero hay una gran cantidad de ellas que por lo general se utilizan para visualizar mejor la información que aparecen en tablas.

En un informe, las recomendaciones que se hacen son: primero va el texto y luego la tabla acompañada de la gráfica, como procedimiento eficaz en el análisis y la interpretación de la información. De esta manera se examinan rápidamente los datos; ade­más, se recuerdan con facilidad, y se aprecian mejor las magnitudes y las regularidades que la variable o variables manifiestan.

Hoy día, se cuentan con una buena herramienta, ya sea en la casa, en la oficina o en la universidad, que permite elaborar en forma rápida una buena representación: el Excel. Esta ofrece diversos tipos de gráficas, ya sea de áreas, barras, columnas, lineales, circulares o pastel, de dispersión, anillos y radar. Existen más tipos de gráficas como: cartogramas, pictogramas, piramidales, las gráficas de Gantt, de escala, entre otras. Al­gunas de ellas de poco uso.

### Gráfica circular o pastel

Recibe este nombre por su forma que permite visualizar los componentes de una totalidad, si se utiliza para mostrar la relación entre las partes y el todo. Cada parte o segmento del pastel representa un elemento o componente de una serie de datos, con la salvedad de que solo puede mostrar una sola serie a la vez.

Es de anotar que los ángulos de cada uno de los sectores son proporcionales a los componentes, siendo más llamativos cuando se colocan colores o con la aplicación de diferentes formas de rayado. La forma de hacer la gráfica consiste en la subdivisión de los 360º de un círculo, proporcionalmente al número o al porcentaje de cada una de las clases o sectores, en que se ha dividido la observación. En la gráfica siguiente, en primer lugar se muestra su aplicación en los atributos, donde con mayor frecuencia se utiliza.

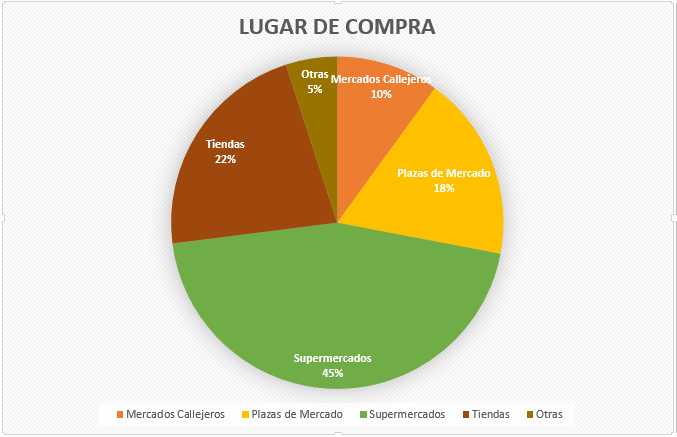
En la mayoría de casos, se utilizan para representar características cualitativas o atributos. En su elaboración se recomienda ordenar los sectores de acuerdo con su magnitud numérica o porcentual; además, hay que tener presente algunos de los inconvenientes que generalmente se observan en su aplicación.

Tabla 11: Distribución porcentual sobre lugares de compra de 400 familias

| Lugar de Compra | % |
| --- | --- |
| Mercados Callejeros | 10 |
| Plazas de Mercado | 18 |
| Supermercados | 45 |
| Tiendas | 22 |
| Otras | 5 |
| Total | 100 |

Distribución porcentual de 400 familias sobre lugares de compra

Gráfica 7: Gráfico Circular de distribución de porcentajes por lugar de compra.



**Descripción Imagen:** Gráfico en el que se muestra un círculo repartido en sectores, que según porcentaje toman una mayor o menor parte de esta, el segmento más pequeño lo tiene otras, la cual equivale al 5%, le sigue Mercados Callejeros que tiene el 10%, le sigue Plazas de mercado 18% luego tiendas 22% y por último Supermercados con el 45%, cada área aumenta proporcionalmente al porcentaje dado.

* Cuando la característica se subdivide en muchas partes, su representación se hace confusa e inconveniente en la visualización de la información, siendo preferible la utilización de otro tipo de figura (por ejemplo, barras).
* Cuando una de las partes que se quiere representar corresponde a una propor­ción o cantidad muy pequeña en relación en las demás; esto hace casi imposible visualizarse, perdiéndose dentro de la gráfica.

### Gráficas de barras o columnas

Se utilizan para comparar elementos individuales o para mostrar cambios fuera de tiem­po; cada barra o columna representa un elemento en una serie de datos. La gran ventaja de estas gráficas es permitir representaciones tanto de características cualitativas como las cuantitativas.

Estas gráficas se comparan de un conjunto de rectángulos (verticales u horizontales) cuya longitud será proporcional a las frecuencias o a las magnitudes correspondientes todas ellas deberán tener el mismo ancho y cuando estas se encuentren separadas, debe­rán tener la misma distancia de separación.

Cuando se utilizan las barras, existen varias formas de representación, ya que puedes ser simples o compuestas y se emplean para representar características cuantitativas cualitativas, como se muestra en la tabla y en las gráficas siguientes.

Valores (miles de millones de $), en las formas de pago en 10 establecimientos comerciales durante un fin semana en la ciudad x

Tabla 12: Valores (Miles de Millones), en las formas de pago en 10 establecimientos comerciales durante un fin de semana en la ciudad x.

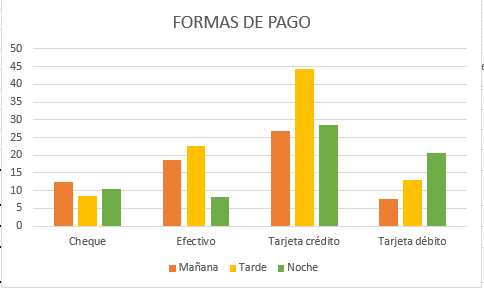
| Formas de pago | Mañana | Tarde | Noche | Total |
| --- | --- | --- | --- | --- |
| Cheque | 12,5 | 8,6 | 10,4 | 31,5 |
| Efectivo | 18,6 | 22,6 | 8,3 | 49,5 |
| Tarjeta crédito | 26,8 | 44,3 | 28,4 | 99,5 |
| Tarjeta débito | 7,6 | 12,9 | 20,6 | 41,1 |
| Total | 65,5 | 88,4 | 67,7 | 221,6 |

### Diagrama de una escala

El término diagrama es muy amplio y puede abarcar varias clases de gráficas. En este caso, son representaciones bastante sencillas en las que se trata generalmente sobre una base común lineal, de longitud proporcional a las cantidades que se desean mostrar. En muchos casos, la gráfica lineal se puede elaborar utilizando rectángulos, sombreando cada parte en forma diferente, donde la longitud total deberá representar al ciento por ciento.

Formas de pago (miles de millones de $) en 10 establecimientos comerciales el fin de semana en la ciudad

Gráfica 8: Diagrama de una escala de Formas de PAGO.



**Descripción Imagen:** Gráfico en el que se 4 grupos de a 3 gráficas de barras según forma de pago, y cada gráfica en grupo corresponde al horario en el que se realiza el pago.

### Gráficas de líneas

Se utilizan para mostrar cambios en los datos a intervalos regulares, casi por lo general correspondiente al tiempo, ya sean horas, días, meses, años. De ahí que se empleen en estadísticas temporales o de tiempo, donde cada línea representa una serie o variable.

En resumen, se podrá afirmar que son líneas que representan una o varias variables, lo cual permite describir los cambios o variaciones que sufren uno o varios hechos, durante un periodo determinado.

En el eje horizontal o abscisa, se señalan los periodos de tiempo y en la línea vertical u ordenada, se tendrán en cuenta los valores o cantidades correspondientes a la variable o variables que se desean representar, cuya intersección permite ubicar un punto y luego unirlos mediante una línea poligonal, reflejando las variaciones que ocurren en esos periodos.

Algunos la denominan como lineal aritmética, dado que ambas escalas que muestran los cambios aritméticos o porcentuales son de igual magnitud; es necesario tener en cuenta que la longitud de ambas líneas sean iguales, ya que una de ellas, al extenderse más que la otra, distorsiona la representación que se quiere hacer.

Tabla 13: Evolución de los precios de tres productos del país: 2007 – 2011

| Periodos | Producto A | Producto B | Producto C |
| --- | --- | --- | --- |
| 2007 | 420 | 380 | 510 |
| 2008 | 210 | 520 | 386 |
| 2009 | 380 | 406 | 612 |
| 2010 | 300 | 708 | 750 |
| 2011 | 605 | 520 | 880 |

Gráfica 9: Representación Lineal

3 polígonos de frecuencias del comportamiento del precio de los productos a lo largo de los años.


**Descripción Imagen:** Gráfico en el que se muestran 3 polígonos de frecuencias una para cada tipo de producto y relacionando el precio registrado por años.

A veces hay necesidad de graficar, de manera diferente de la anterior, pues existen otras gráficas, como: la semilogarítmica, cuando una de las variables en vez de presentir un crecimiento aritmético lo hace en forma geométrica, por tanto, ese eje debe estar dada en escala logarítmica, mientras que la otra se da cii forma aritmética. De ahí que sea necesario la utilización del papel semilogarítmico; la logarítmica, cuyas dos variables presentan crecimientos geométricos. Para ello es necesario utilizar el papel logarítmico.

### Pirámide de población

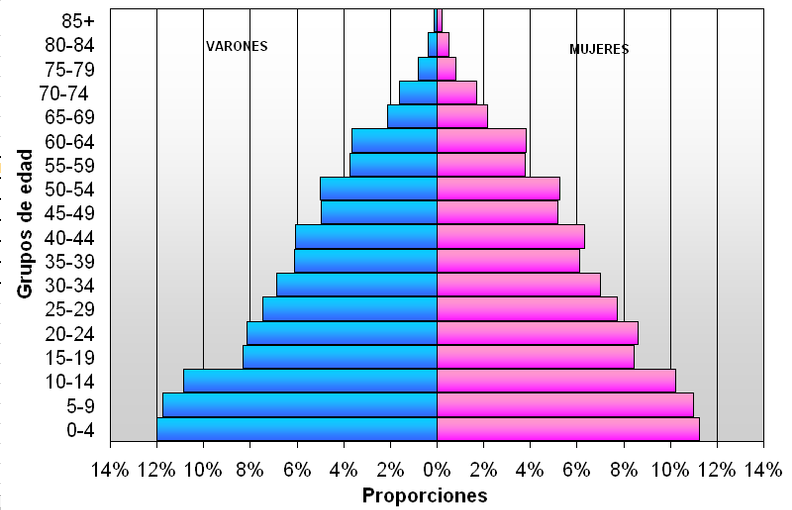
Nombre que se le da a aquella representación gráfica en la cual la población se encuen­tra, por ejemplo, clasificada de acuerdo con su edad (variable) y sexo (atributo); se utiliza, preferentemente, en representaciones de estadísticas mixtas de investigaciones demográficas, económicas y sociales.

Esta gráfica no suele ser más que un histograma, el cual permite representar dos características en una misma gráfica.

Tabla 14: Matrícula por sexo y cursos

| GRADOS | FEMENINO | MASCULINO |
| --- | --- | --- |
| Primero | 480 | 320 |
| Segundo | 420 | 312 |
| Tercero | 400 | 308 |
| Cuarto | 396 | 300 |
| Quinto | 310 | 250 |
| Sexto | 300 | 220 |
| Séptimo | 280 | 210 |
| Octavo | 276 | 198 |
| Noveno | 250 | 180 |
| Décimo | 200 | 150 |

Gráfica 10: Pirámide Poblacional.



**Descripción Imagen:** Gráfico de barras horizontales una contigua a la otra, un grupo de estos con orientación a la derecha y el otro con orientación a la izquierda, según frecuencia por género, el eje y, son los grupos de edades y empieza de 0 has 85 y más.

### Pictogramas

Se consideran la representación por medio de figuras que guardan relación con la característica que se desea visualizar. Es un método restringido a la representación de relacio­nes simples, con el fin de eliminar la comparación entre objetos de tamaños diferentes, pues en la mayoría de los casos distorsionan la realidad.

Se acostumbra representar un valor unitario por un símbolo, el cual describe la magnitud en cuestión, repitiendo ese símbolo típico un número adecuado de veces hasta alcanzar el valor total. Sin embargo, es necesario colocar a continuación la cantidad o el valor total que se va a representar, acompañada de una figura con los valores dados, indicando el valor correspondiente a cada una de ellas.

Siempre que elabore un pictograma, se debe tener presente que esa figura explique, por sí misma, la característica que se quiere representar; se aconseja no utilizar dimensión, volumen o tamaño del símbolo. Además, se debe procurar dibujar únicamente el número apropiado de símbolos o figuras, dependiendo de la cantidad que se va a representar, dando de esta manera una visión global del problema, para evitar expresar en detalle el comportamiento del hecho.

También se pueden utilizar figuras de objetos, personas o animales, de igual tamaño y con el mismo valor, colocando siempre tanto el valor total y el de cada objeto.

Finalmente, cuando se tenga la oportunidad de leer periódicos o revistas y se vean estas representaciones gráficas, hay que ser cuidadosos, pues pueden haber utilizado equívocos que desorientan al interpretarlas. Vale la pena recordar que los dibujos, en la mayoría de los casos, dan vida al informe pero también pueden ocasionar, en algunos casos, una falsa impresión, la cual debe evitarse cuando se decide usarlas.

**Gráfica 11:** Pictogramas con el uso de figuras



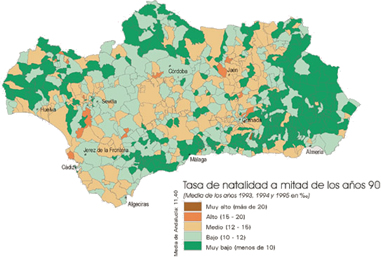
**Descripción Imagen:** símbolos de prohibición de algunas actividades, expresadas mediante el circulo atravesado por un línea, prohibición de fumar, de apagar con agua de tocar, entre otras.

### Cartogramas

Se consideran muy expresivos porque permiten deducir fácilmente las características de cada región o de todo el país, así como las diferencias que puedan existir entre ellas.

Generalmente se emplean para visualizar información estadística de carácter descriptivo, por medio de mapas que contienen figuras, símbolos o cualquier otro elemento correspondiente a la información. Por ejemplo, los que se observan, todos los días, en televisión o en los periódicos y que muestran el estado del tiempo o los lugares del país en conflictos.

Gráfica 12: Cartograma de tasa de natalidad.



**Descripción Imagen:** Mapa de una zona geográfica con algunas zonas marcadas con diferentes colores según la escala registrada en cuanto a tasa de natalidad de esta.

### Gráficas de Gantt

En muchas empresas, es indispensable la utilización de esta valiosa herramienta gráfica, en las que se establecen las diferentes etapas de trabajo por ejecutar y al mismo tiempo, durante ese periodo establecido, van señalándose la cantidad realizada, lo cual permite saber si se está cumpliendo lo establecido o, por el contrario, alguna o algunas de ellas se encuentran retrasadas.

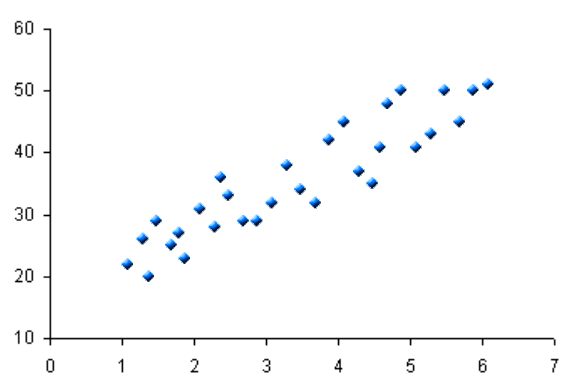
Hoy día, se ha modernizado su presentación y utilización y se distinguen varias clases de gráficas de Gantt.

* Gráfica de avance del trabajo: permite visualizar la forma como se avanza en la. ejecución del trabajo, basado en un plan; en algunos casos, esto permite trasladar recursos a aquellos que presentan retardos o demasiado avance en la ejecución.
* Gráfica de trabajo de los obreros y máquinas: como su nombre lo indica, permite hacer control por obrero o por máquina, en todo el proceso de desarrollo tanto en el volumen de trabajo, como en el aprovechamiento del tiempo que se ha indicado que debe llevarse a cabo en un periodo determinado.
* Gráfica del programa de trabajo: visualiza el tiempo utilizado y la tarea por realizar, haciendo de esta manera mucho más equitativo la ejecución de los diferentes procesos, evitando la presentación de algunas inactividades (tiempo y recurso), en el desarrollo de alguna operación o en el trabajo en general.

### Diagrama de dispersión

Más conocido como nube de puntos, se usa para mostrar la relación que puede ha: entre dos variables o series de datos. Se aplica mucho en el análisis de regresión y correlación, que se verá más adelante, como una forma práctica para determinar y analizar la línea que mejor se ajusta a un conjunto de puntos.

**Gráfica 13:** Diagrama de dispersión



**Descripción Imagen:** Primer cuadrante del plano cartesiano sobre el cual se encuentran registrados distintos puntos, estos relacionan un valor en X con uno en Y.

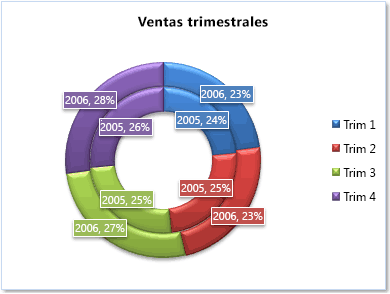
### Gráfica en forma de anillo

En algunos casos, se utiliza para mostrar la relación entre las partes y el todo. Es diferen­te del gráfico circular, pues permite visualizar más de una serie de datos, dado que cada anillo representa un conjunto de información.

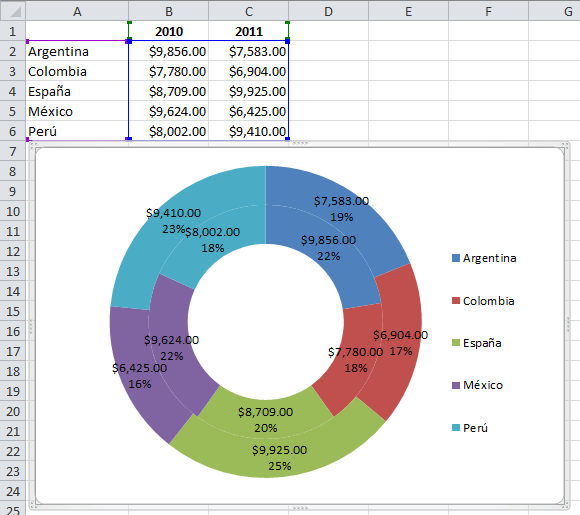
### Gráfica en forma de radial

Con poca frecuencia, se utiliza para comparar elementos en algunas series de datos, en la que cada una muestra una línea que se ubica alrededor de un punto central, que indica el comportamiento de cada serie de datos.

Gráfica 14: Gráfica en forma de anillo

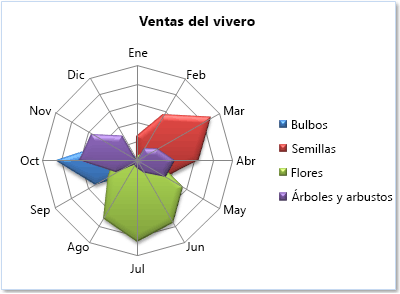


**Descripción Imagen:** Dos anillo concéntrico, uno de radio mayor, contiguo a uno de radio menor sobre cada uno se encuentra distribuido el porcentaje por trimestre en un año en particular.



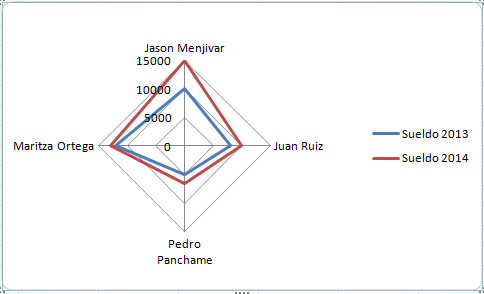
**Descripción Imagen:** Dos anillo concéntricos, uno de radio mayor, contiguo a uno de radio menor sobre cada uno se encuentra distribuido el porcentaje por país en un año en particular.

**Gráfica 15:** Gráfico radial de relación de tipo de planta y ventas.



**Descripción Imagen:** Plano cartesiano sobre el cual se encuentran decágonos cada vértice representa un mes del año y se relacionan las ventas por tipo de planta.

**Gráfica 16:** Gráfico radial de relación de salarios.



**Descripción Imagen:** Plano cartesiano sobre el cual se encuentran cuadrado con vértices sobre los ejes, cada eje representa una persona y su ingreso salarial por año.

### Eterogramas

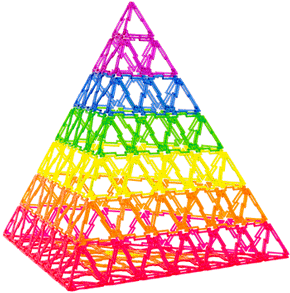
Se puede considerar como un conjunto de dos pruebas fotográficas o los dibujos desti­nados a observarse, a través de un estereoscopio; es una representación tridimensional de una figura sobre una superficie plana.

### Gráficas que utilizan cuadrados y triángulos

Como su nombre lo indica, son gráficas en cuya representación se utilizan cuadrados y triángulos. Su elaboración requiere cierta habilidad para determinar el área de forma correcta, superponiendo las figuras en vez de yuxtaponerlas. De todas formas, la visualización de la información no es tan clara, como debe ser el objetivo de una graficación.

**Formas de triángulos y cuadrados**

Gráfica 17: Forma de triángulo.



**Descripción Imagen:** Pirámide triangular con cambio de color en su altura, dada la evolución del proceso que describa.

Gráfica 18: Cubo de colores.



**Descripción Imagen:** Imagen de un cubo el cual contiene distintos colores sobre cada una de sus caras.

## ALGUNOS COMENTARIOS ACERCA DE LAS GRÁFICAS

Las gráficas son una forma de visualizar mejor la información, pero esta nunca sustituye en la tabla que contiene los datos, solo la complementa. En un informe con una presentación adecuada, primero debe ir el texto, que incluye la explicación de los resultados; luego aparece la tabla con la información correspondiente; por último, se anexa la gráfica que permita visualizar mejor la información que se desea presentar.

La apariencia de la gráfica depende de la escala usada y de los datos presentados.

Puede ocurrir que una misma línea tenga una significación diferente, según la posición en cuanto a la relación con la magnitud de los valores presentados. Estos se pueden observar en los efectos de la publicidad, cuya ilustración muchas veces produce un gran efecto sobre los hechos mismos. Basta cambiar la escala del dibujo o exagerando la significación aparente de la línea, para lograr el efecto deseado.

Quizá no sea exagerado afirmar que de todos los métodos estadísticos, ninguno es más útil para engañar a incautos que la gráfica estadística.

En un informe se recomienda si existe más de una gráfica, numerarlas; y cada una de ellas llevar el título correspondiente que explique claramente su contenido; es indispen­sable el uso de convenciones (rayado, colores, etcétera) que identifique fácilmente cada variable por representar. La escala debe ser bastante amplia, de tal forma que permita revelar diferencias o tendencias significativas, ya que el ojo del lector no distingue fácilmente ligeras diferencias, ni debe ampliarse hasta dar énfasis inadecuado a detalles significantes; por tanto, deberá conceder atención a los valores que toman las escalas a fin de no distorsionar aquello que deberá representarse.

### Practica lo aprendido

1. Con los siguientes datos, correspondientes a una variable continua, elabore una tabla de frecuencias y las gráficas de polígono y ojiva.

46 39 64 58 72 77 52 76 43 46 78 73 61 68 42 54 82 70 63 59 49 80 40 49 72 62 58 66 55 60

1. Con los siguientes datos, el gerente de la compañía le propone hacer dos (2) gráficas con el fin de escoger una de ellas para la próxima publicación de la revista de la empresa.

Tabla 15: Datos para publicación.

| Años | Producción | Ventas |
| --- | --- | --- |
| 2006 | 380 | 760 |
| 2007 | 450 | 680 |
| 2008 | 780 | 800 |
| 2009 | 620 | 905 |
| 2010 | 930 | 1500 |
| 2011 | 1080 | 1300 |

1. En una pequeña encuesta realizada por las autoridades del barrio, sobre el número de personas que trabajan en industrias caseras y que emplean menores de edad, se obtuvo la siguiente información:

Tabla 16: Cursos por trabajador.

| TRABAJADOR | CURSOS |
| --- | --- |
| 1 | 5 |
| 2 | 6 |
| 3 | 5 |
| 4 | 2 |
| 5 | 4 |
| 6 | 2 |
| 7 | 4 |
| 8 | 1 |
| 9 | 5 |
| 10 | 3 |
| 11 | 3 |
| 12 | 5 |
| 13 | 5 |
| 14 | 2 |
| 15 | 4 |
| 16 | 6 |
| 17 | 8 |
| 18 | 10 |
| 19 | 12 |
| 20 | 14 |
| 21 | 16 |
| 22 | 18 |
| 23 | 20 |
| 24 | 22 |
| 25 | 24 |

Como ayuda en la lectura de la anterior información presentada, los datos se organizaron en una tabla de frecuencias. Realice la gráfica de barras que representa la anterior situación.

1. Busque un ejemplo que se adapte a las condiciones y complete y responda:
2. El número de establecimientos cuyas ventas diarias fluctuaron entre $15  
   millones y $45 millones es igual a:
3. El porcentaje (%) de establecimientos cuyas ventas diarias fue de por lo menos  
   $25 millones es igual a:
4. El número de establecimientos cuyas ventas diarias fueron superiores a $35  
   millones es igual a:
5. El porcentaje (%) de establecimientos cuyas ventas diarias se encuentran entre  
   $5 y $15 millones es igual a:
6. Construya el polígono de frecuencias de la distribución. ¿Qué concluye?
7. Construya una tabla de frecuencias.
8. Cree una tabla que se adapte a las condiciones y complete y responda:
9. El porcentaje (%) de clientes habituales es igual a:
10. Entre los clientes esporádicos, el porcentaje (%) de los que pagan en efectivo es igual a:
11. Entre los clientes habituales, el porcentaje (%) de los que pagan con tarjeta de crédito es igual a:
12. El número de clientes que pagan con tarjeta de crédito es igual a:
13. Elabore un cuadro o tabla con esta información.
14. Una distribución de frecuencias relativas representa las frecuencias en términos de
15. Números enteros.
16. fracciones.
17. porcentajes.
18. Las opciones (a) y (b).
19. Todas las anteriores.
20. Histograma y la ojiva se utilizan para representar:
21. distribuciones cualitativas,
22. variables discretas.
23. variables continuas.
24. ambas (b) y (c), pero no (a).
25. ninguna de las anteriores.
26. La fórmula de Sturges 1 + 3,3 log n, se utiliza para determinar:
27. Frecuencias absolutas
28. Frecuencias relativas
29. Número de intervalos
30. Amplitud del intervalo.
31. marcas de clase.
32. Busque un gráfico, sobre el porcentaje de estudiantes, de acuerdo con el número ­de libros leídos, en el último semestre del 2011, asuma que la muestra fue realizada a 500 estudiantes de la jornada diurna en nuestra universidad.
33. Elabore una tabla de frecuencias.
34. ¿Qué porcentaje de estudiantes leyeron menos de 6 libros?
35. ¿Cuántos estudiantes leyeron más de 5 libros?
36. ¿Qué nombre recibe la gráfica?
37. Señale como mínimo cinco (5) recomendaciones que se deben tener en la elaboración de una buena gráfica.
38. Como regla general, ¿qué cantidad de intervalos o marcas de clase se recomienda para organizar datos en una tabla de frecuencias?
39. Menos de cinco.
40. Más de 25.
41. Entre 20 y 25.
42. Entre 5 y 16.
43. Ninguna de las anteriores.
44. Una corporación de ahorro y vivienda desea conocer el número de deudores en el país con más de 6 meses de vencido el plazo, al finalizar el año 2011.

Los departamentos de crédito de cada sucursal informaron lo siguiente.

Tabla 17: Deudores morosos por sucursal.

| Número de Deudores Morosos | Acumulado número de sucursales |
| --- | --- |
| 52 | 8 |
| 108 | 12 |
| 120 | 20 |
| 210 | 36 |
| 306 | 54 |
| 510 | 80 |

1. Complete la tabla de frecuencias.
2. Calcule el porcentaje de sucursales con menos de 210 deudores morosos.
3. ¿Qué porcentaje de sucursales tienen más de 108 y menos de 306 morosos?
4. Elabore una gráfica con los anteriores datos.
5. En la construcción de una distribución de frecuencias, al dividir el valor del rango por el número de intervalos, el resultado obtenido indica:
6. las frecuencias absolutas.
7. el valor del recorrido.
8. la amplitud del intervalo.
9. los límites del intervalo.
10. todas las anteriores.
11. Realice una encuesta entre 500 estudiantes sobre la distribución del tiempo, cuando se termina el tiempo de estudio en la universidad, realice una gráfica y:

Complete:

1. El porcentaje de los estudiantes que prefieren ver Tv. y lecturas practican un deporte y comparten su tiempo es:
2. El número de estudiantes que se dedican a estudiar, investigar y hacer tarea es:
3. El número de estudiantes que no realizan ninguna actividad es:
4. Elabore una tabla de frecuencias.
5. Quinientas (500) empresas de un sector de la economía distribuye utilidades (en Bullones de $) entre sus accionistas, como se muestra en la tabla:
6. ¿Cuántas empresas tienen utilidades entre $30 y $150 millones?
7. ¿Cuántas empresas tienen pérdidas superiores a los $ 10 millones?
8. ¿Cuántas empresas tienen utilidades superiores a los $120 millones?
9. Elabore una gráfica correspondiente a las utilidades recibidas por las 500 empresas.

Tabla 18: Rango de Utilidades y Porcentajes de Utilidades

| Utilidades | Porcentajes |
| --- | --- |
| -20 – -10 | 6% |
| -10 – 10 | 10% |
| 10 – 30 | 14% |
| 30 – 40 | 24% |
| 40 – 60 | 20% |
| 60 – 90 | 10% |
| 90 – 150 | 8% |
| 150 – 170 | 8% |

1. Una encuesta realizada a 300 estudiantes respecto a consultas hechas en la biblioteca de la institución, según los temas requeridos, encontró que 86 de ellas eran de economía, 50 de matemáticas, 70 de derecho, 26 de inglés, 60 del área contable y 8 de diferentes temas. Elabore una tabla de frecuencias con la respectiva gráfica.
2. Suponga que se realiza una investigación sobre precios (miles de $) de un artículo distribuido en la ciudad, con los siguientes resultados:

250 265 250 290 238 260 248 270 232 275 262 262 240 270 272 236 263 258 248 254 280 263 246 236 288 254 270 263 265 240 270 280 260 250 246 260 284 290 272 268 230 300 284 240 254 258 296 278 280 278.

1. Elabore una tabla de frecuencias, trabajando con una amplitud constante. Obtenga el número de intervalos aplicando m = 1 + 3,3log n.
2. Dibuje el histograma, el polígono y la ojiva.
3. ¿Aproximadamente qué porcentaje de establecimientos vende el producto a j» precio inferior a $270?
4. La sección de cobro de una empresa comercial registra los días de mora en el cum­plimiento de sus obligaciones, en un mes, a 64 deudores. Los resultados son:

0, 4, 4, 20, 3, 5, 0, 3, 3, 4, 2, 0, 10, 5, 3, 3, 0, 0, 0, 2, 0, 2, 0, 10, 5, 2, 8, 0, 5, 3, 5, 10, 3, 5, 10, 5, 0, 4, 4, 0, 20, 10, 0, 10, 16, 10 12, 25, 12, 2, 16, 8, 12, 3, 2, 3, 5, 3, 2, 16, 5, 25, 8, 4.

1. Considere los días como variable discreta y elabore una tabla de frecuencia.
2. Dibuje el diagrama de frecuencias correspondiente.
3. La empresa da tres días de gracia, en los cuales no se cobra intereses por mora. ¿Qué proporción de deudores tendrán que pagarlos?
4. Explique qué diferencia puede haber entre una gráfica de barra y un histograma.
5. Una empresa realiza una encuesta a 800 empleados respecto al traslado de sus cesantías a los fondos de pensión, recientemente creados. El objetivo es determinar cuántos empleados están dispuestos a aceptarlo. Los resultados obtenidos son:

Tabla 19: Votación.

| SI | NO | INDECISOS |
| --- | --- | --- |
| 256 | 474 | 70 |

1. Convierta los datos en porcentajes y construya
2. un diagrama circular.
3. un diagrama de barras.
4. un diagrama de barras de componentes en porcentajes (% acumulado horizontal).
5. Una muestra realizada a 250 familias de un barrio de la ciudad, sobre el consumo de chocolate en el desayuno, de acuerdo con sus preferencias por la marca del producto arrojó los siguientes resultados:

Tabla 20: Número de familias por marca.

| Marca | Número de Familias |
| --- | --- |
| A | 116 |
| B | 44 |
| C | 182 |
| D | 80 |
| E | 58 |

Construya la gráfica que mejor representa a este conjunto de observaciones. Explique el porqué de su decisión.

1. Responda:
2. ¿Cuándo se aconseja usar la distribución de frecuencias para organizar datos?
3. ¿En qué casos se debe granear en papel semilogarítmico?
4. ¿Cuándo en papel logarítmico?
5. Según la regla de Sturges m = 1 + 3,3log n, ¿cuántas clases o marcas de clases de obtiene:

* ¿Si la muestra es de 80 observaciones?
* ¿Si la muestra es de 150?

1. Realice una pequeña encuesta con sus compañeros de clases, solicitándoles información sobre: edad, peso, estatura, promedio de calificación en una asignatura, número de libros consultados en la biblioteca, durante un determinado periodo. Organice los datos obtenidos y elabore las gráficas correspondientes.

# BIBLIOGRAFÍA

Martínez, Ciro. – Levin, R. (2012). *Estadística Aplicada*, primera edición. Colombia: Pearson.