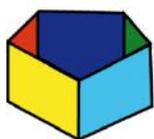


# Reforma de la Educación Matemática en Costa Rica



## Curso bimodal para el Tercer Ciclo: Enfoque de Resolución de problemas



Imagen cortesía de ddpavumba y FreeDigitalPhotos.net

## Estadística 2011

---

## Tabla de contenido

Presentación.....	1
Introducción.....	3
I. Generalidades acerca de la enseñanza de la Estadística y la Probabilidad .....	4
Actividad 1 .....	4
Análisis de la Actividad 1.....	4
II. Nociones básicas de Estadística.....	7
Actividad 2 .....	7
Análisis de la actividad 2 .....	7
Actividad 3 .....	8
Análisis de la Actividad 3.....	9
Actividad 4 .....	11
Análisis de la Actividad 4.....	12
Actividad 5 .....	13
Análisis de la Actividad 5.....	14
Actividad 6 .....	15
Análisis de la Actividad 6.....	16
Actividad 7 .....	19
Análisis de la Actividad 7.....	19
III. Recolección de información .....	22
Actividad 1 .....	22
Análisis de la Actividad 1.....	22
Actividad 2 .....	24
Análisis de la Actividad 2.....	24
Actividad 3 .....	26
Análisis de la Actividad 3.....	27
Actividad 4 .....	27
Análisis de la Actividad 4.....	28
Actividad 5 .....	29
Análisis de la Actividad 5.....	30
Actividad 6 .....	31
Análisis de la Actividad 6.....	34
IV. Representaciones .....	37
Actividad 1 .....	37
Análisis de la Actividad 1.....	38
Actividad 2 .....	39
Análisis de la Actividad 2.....	39
Actividad 3 .....	40
Análisis de la Actividad 3.....	42
Actividad 4 .....	48
Análisis de la Actividad 4.....	49
Actividad 5 .....	50
Análisis de la Actividad 5.....	50
Actividad 6 .....	51
Análisis de la Actividad 6.....	51

Actividad 7 .....	53
Análisis de la Actividad 7.....	53
Actividad 8 .....	54
Análisis de la Actividad 8.....	55
Actividad 9 .....	55
Análisis de la Actividad 9.....	55
Actividad 10 .....	57
Análisis de la Actividad 10.....	58
Actividad 11 .....	60
Análisis de la Actividad 11.....	60
Actividad 12 .....	61
Análisis de la Actividad 12.....	62
Actividad 13 .....	63
Análisis de la actividad 13 .....	63
Actividad 14 .....	65
Análisis de la Actividad 14.....	65
Actividad 15 .....	67
Análisis de la Actividad 15.....	68
V. Medidas resumen de posición y de variabilidad .....	69
Actividad 1 .....	69
Análisis de la Actividad 1.....	70
Actividad 2 .....	71
Análisis de la Actividad 2.....	72
VI. Recomendaciones metodológicas.....	73
Propuesta de un problema .....	73
Trabajo estudiantil independiente .....	74
Discusión interactiva y comunicativa.....	75
Clausura o cierre .....	76
Bibliografía.....	77
Créditos .....	78



## Presentación

El *Curso bimodal para el Tercer Ciclo: Enfoque de Resolución de problemas* forma parte del proyecto *Reforma de la Educación Matemática en Costa Rica*. Este proyecto del Ministerio de Educación Pública es apoyado por la Fundación Costa Rica-Estados Unidos de América para la Cooperación.

Este proyecto ha buscado y buscará apoyar la reforma de la educación matemática en Costa Rica por medio de la elaboración de un nuevo currículo escolar y de documentos de apoyo curricular, la capacitación de docentes y la creación de medios que apoyen la implementación de los programas, objetivos macro a realizar con base en prácticas exitosas en la enseñanza de las Matemáticas y resultados positivos de la investigación tanto a nivel nacional como internacional. La población con la que este proyecto trabaja directamente son educadores de primaria y secundaria que deben enseñar Matemáticas, asesores pedagógicos y nacionales, y otros funcionarios del MEP.

Este proyecto cobra gran trascendencia luego de conocerse en el 2011 los resultados en el rendimiento de Costa Rica en las pruebas PISA 2009+, que revelan que el país posee importantes debilidades en Matemáticas. El progreso nacional obliga a medidas de gran envergadura para poder responder con seriedad a esta realidad. Este proyecto ofrece una respuesta integral a los desafíos colocados por este diagnóstico ineludible de tomar en cuenta.

El curso bimodal para el Tercer Ciclo posee como objetivo familiarizar a los docentes con el enfoque principal de los nuevos programas de estudio: la resolución de problemas, con especial énfasis en contextos reales. Para ello incluye dos tipos de unidades didácticas: el primero busca aportar elementos de la fundamentación del currículo, y el segundo presentar varias situaciones educativas en las diversas áreas matemáticas de este ciclo mediante las cuales se pueda trabajar con ese enfoque. Dominar los principales elementos de la fundamentación general es indispensable para poder comprender y llevar a las aulas con efectividad los nuevos programas. Es por eso que se solicita a los participantes de este curso comenzar con una amplia dedicación a su estudio y a la realización de las prácticas que se incluyen. Solo así será posible visualizar y manejar con propiedad las otras unidades. No obstante, se da flexibilidad al participante para realizar las prácticas a lo largo de todo el curso.

Se ha decidido, en cuanto al segundo tipo de unidades, abarcar cuatro de las áreas que se proponen en los nuevos programas. En *Números* se deja de lado la visión conjuntista para dar paso al tratamiento del sentido numérico, el cálculo operacional y mental. *Geometría* que incluye tópicos relacionados a la visualización espacial, transformaciones geométricas y plano cartesiano. *Estadística y Probabilidad* aunque sí se contemplaba en los programas anteriores, no existía un trabajo continuo y articulado de los conceptos estadísticos y de probabilidad como el que se ofrece ahora desde primaria. *Relaciones y Álgebra* como novedad introduce el trabajo con sucesiones y el tratamiento de la función lineal y la cuadrática. El área de *Medidas* se trabajará de forma transversal respecto a las otras áreas antes mencionadas. Estas cuatro unidades poseen una gran unidad que se la brinda el propósito de todo el curso: comprender y usar el enfoque del currículo. No todos los tópicos del Tercer Ciclo se incluirán en este curso, solo algunos que son más novedosos o que se prestan mejor para mostrar el enfoque. Es decir, este curso no pretende ofrecer una capacitación completa. Se busca dar algunos elementos al docente para que éste en el desarrollo de su acción

profesional autónoma siga ampliando su dominio del enfoque curricular, de los contenidos programáticos y de la forma de trabajarlos en las aulas.

En la elaboración de esta unidad han participado diversas personas como autores, revisores, editores temáticos y de estilo y forma y varios colaboradores. Ha sido producto de un amplio esfuerzo colectivo realizado con mucha seriedad y profesionalismo, con mucho cariño y con ritmos de tiempo muy intensos.

En el 2013, sin embargo, se desarrollarán otros cursos bimodales en esencia con los mismos propósitos, pero esta vez enfatizando algunas dimensiones incluidas en los programas, como el uso de la historia de las matemáticas y el uso de las tecnologías. En el 2014, otros cursos bimodales brindarán mayor atención a la Estadística y Probabilidad.

A partir del 2013 se aportarán cursos totalmente virtuales que permitirán repetir los cursos bimodales con otra modalidad, y reforzar los medios para ampliar la capacitación a más educadores.

A partir del 2013 también se contará con una comunidad virtual especializada para la educación matemática que permitirá integrar varias de las diversas acciones de capacitación y de implementación de los programas, y servir como un medio dinámico para compartir experiencias y para obtener recursos didácticos.

Para la implementación eficaz de los nuevos programas y para avanzar en la reforma de la Educación Matemática en el país, se está diseñando este año un plan de transición, y también se llevarán a cabo planes pilotos en la Primaria y Secundaria del 2012 al 2014.

Todas estas acciones poseen un efecto integrador y sinérgico.

Deseamos que este curso pueda resultarles de gran provecho y sobre todo de motivación para avanzar en los cambios que en la enseñanza y aprendizaje de las matemáticas requieren nuestros niños y jóvenes.

Cordialmente,

**Ángel Ruiz**

Director general

Proyecto *Reforma de la Educación Matemática en Costa Rica*.

# Estadística



## Introducción

Los diversos medios de comunicación juegan un rol cada vez más importante en la sociedad, aquel ciudadano que posea dominio sobre la información que le rodea tendrá una mejor oportunidad de desarrollarse dentro de la misma.

Los temas de Estadística y Probabilidad se incluyen desde hace varios años en el currículo tanto de primaria como de secundaria de muchos países.

Debido al carácter instrumental de la Estadística, para otras disciplinas y la importancia de su razonamiento dentro de una sociedad caracterizada por la disposición de información, la necesidad de analizarla y tomar decisiones basadas en un conjunto de datos, es que se le ha dado importancia a su enseñanza.

Este material está dirigido a docentes de Matemática de secundaria, pretende ser un apoyo para el educador en cuanto a conceptos y tipos de actividades que puede realizar al abordar el tema de Estadística y Probabilidad en el III ciclo de la Educación Obligatoria costarricense.

# I. Generalidades acerca de la enseñanza de la Estadística y la Probabilidad

## Actividad 1

Reflexione sobre las siguientes preguntas y encuentre posibles respuestas:

1. *¿Qué es la Estadística?*
2. *¿Para qué sirve la Estadística?*
3. *¿En qué áreas del conocimiento o campos laborales se utiliza Estadística?*
4. *¿Por qué la Estadística se puede vincular con otras áreas del conocimiento?*
5. *¿Qué son datos?*
6. *¿Para qué sirven los datos?*
7. *Mencione situaciones en las que se requiere utilizar datos.*
8. *Indique algunas situaciones cotidianas en las que se utilizan conceptos estadísticos.*
9. *¿Qué es la Probabilidad?*
10. *¿Para qué sirve la Probabilidad?*
11. *¿En qué áreas del conocimiento o campos laborales se utiliza Probabilidad?*
12. *Indique algunas situaciones cotidianas en las que se utilizan conceptos probabilísticos.*
13. *¿Cuál es la importancia de generar aprendizaje en conceptos de Estadística y Probabilidades desde los primeros años de escolaridad?*
14. *¿Qué habilidades y destrezas en Estadística y Probabilidades piensa usted que requiere un ciudadano?*

## Análisis de la Actividad 1

La reflexión sobre las preguntas anteriores es fundamental si se desea orientar un proceso de enseñanza y aprendizaje de la Estadística y la Probabilidad con jóvenes. Específicamente en Estadística, se debe tener claro que esta disciplina es mucho más que simples técnicas y métodos para recolectar, resumir o presentar información, sino que es una ciencia que analiza la forma en que la información debe ser utilizada para orientar la comprensión del entorno del que provienen los datos y, por ende, facilitar la argumentación y la toma de decisiones con base en esa información.

Debido al rol de la Estadística como disciplina fundamental en la utilización de la información para la comprensión del entorno, su enseñanza ha tomado auge a nivel mundial, y está incluida en los currículos educativos de primaria y secundaria de la mayoría de países.

En Costa Rica, la Estadística se incluyó dentro del programa de estudio de los tres ciclos de la Educación General Básica desde 1995; sin embargo, son pocos los docentes que desarrollan estas temáticas en sus clases y en otras ocasiones los

imparten de manera superficial, ya sea por falta de tiempo, para dar mayor énfasis en otros temas matemáticos que consideran más importantes, o porque no han recibido una adecuada formación para enseñar esta disciplina.

Consientes que la Enseñanza de la Matemática es un pilar fundamental en la educación del ciudadano y que el área de Estadística es una poderosa herramienta para la resolución de problemas cotidianos, el nuevo currículo educativo de matemáticas incluye un mayor énfasis en esta disciplina.

Respecto a la Probabilidad, según Batanero (2001), la intuición tiene un papel determinante, dado que los modelos intuitivos tienen dos grandes funciones: desde edades tempranas permite que el niño comprenda el entorno por sus propios medios antes de ser capaz de entender la complejidad del modelo matemático y además, prepara el conocimiento analítico que tendrá que emplear posteriormente.

Al terminar la instrucción básica, el estudiante debe ser un “consumidor” inteligente y conocer suficiente acerca del origen de los datos y el tipo de razonamiento utilizado en el análisis de los mismos; ya sea que el individuo los obtenga de manera personal o los tome de alguna fuente en particular.

En esta sección se analiza una serie de conceptos elementales vinculados con la Estadística y la Probabilidad, cuya comprensión viene a favorecer la interpretación de las situaciones de aprendizaje que se desarrollarán a lo largo del documento.

Debido a que en el ámbito preuniversitario no se pretende formar estadísticos profesionales, sino potenciar una cultura estadística y probabilística que le permita al joven aprovechar la información que se genera en el entorno para favorecer el análisis, la argumentación y la toma de decisiones, los programas vigentes únicamente incluyen elementos de Estadística descriptiva. Por esta razón, los diferentes tópicos analizados en el documento se relacionan con dicha área.

\*\*\*

A continuación se citan las habilidades generales que se espera que desarrolle el lector con esta unidad didáctica y los conceptos que se tratan en ella.

## **Habilidades generales**

- Comprender los conceptos de Estadística: datos, población, muestra, variables cuantitativas (discretas y continuas), variables cualitativas (nominales y ordinales).
- Conocer las distintas representaciones tabulares y gráficas y sus principales características.
- Utilizar las representaciones gráficas y tabulares para resumir información que permita una mejor interpretación de los fenómenos.
- Conocer las diversas técnicas para la recolección de datos existentes y no existentes.
- Recolectar datos del entorno por medio de observación, experimentación e interrogación.
- Combinar diversas técnicas para recolección de datos que sean insumos para el análisis de información relacionado con problemas concretos.
- Valorar la importancia de las medidas de resumen: posición, tendencia central y variabilidad para el análisis de la información estadística.

- Utilizar las medidas de posición para resumir y analizar la información proveniente de un conjunto de datos cuantitativos.
- Utilizar las principales medidas de variabilidad para evaluar y comparar la dispersión de los datos.
- Analizar la importancia del uso medidas relativas de tendencia central y variabilidad dentro de los análisis comparativos de información.

## **Conocimientos**

- Nociones básicas de Estadística (población, muestra, unidad estadística, datos, observación)
- Tipos de variables; cualitativas (nominales y ordinales) y cuantitativas (discretas y continuas).
- Representaciones tabulares (distribuciones de frecuencia, cuadros de datos cruzados).
- Representaciones gráficas (gráficas lineales, circulares, barras simples y comparativas, histogramas, polígonos de frecuencias y diagrama de puntos).
- Recolección de información (observación, interrogación abierta y cerrada, experimentación, cuestionario).
- Técnicas de recolección de información existente y no existente.
- Fuentes primarias y secundarias.
- Medidas de posición o localización (máximo, mínimo, moda, media, mediana, cuartiles).
- Distribuciones de datos simétricas y asimétricas.
- Diagramas de cajas.
- Medición de la variabilidad de los datos (recorrido, recorrido intercuartílico, variancia, desviación estándar).
- Estandarización o tipificación de mediciones.
- Variabilidad relativa: coeficiente de variación.
- Promedio ponderado.

## II. Nociones básicas de Estadística

### Actividad 2

La noticia siguiente fue publicada en el periódico La Nación del 10 de mayo del 2011:

#### EE. UU.: 7,5 millones de usuarios de Facebook tienen menos de 13 años

**Washington (AFP).** De los 20 millones de menores usuarios de Facebook en Estados Unidos, 7,5 millones tienen menos de 13 años, de los cuales un millón han sido intimidados, hostigados o amenazados a través de esta red social, reveló un estudio publicado el martes.

El año pasado, más de 5 millones de usuarios de Facebook tenían 10 años o menos. En general, fueron autorizados a usar la red social sin haber sido vigilados por sus padres, según el estudio del grupo de defensa de consumidores Consumer reports. Las amenazas a las que se exponen estos niños van desde deseos malintencionados hasta acoso sexual, expresa el informe.

Facebook pide a sus usuarios tener al menos 13 años para poder inscribirse. Pero muchos niños, con o sin la complicidad de sus padres, se abren cuentas con una fecha de nacimiento falsa.

Los padres de los niños de 10 años o menos "parecen desentenderse bastante" del uso que sus hijos le dan a la red social, porque piensan que un niño está menos expuesto a amenazas en Internet que un adolescente, indica el estudio.

Pero mientras un niño de 10 años podría no interesarse en descargar pornografía de la red, "necesita una protección frente a los riesgos que trae Internet, como las invitaciones enviadas por adultos desconocidos", destaca el informe.

Más de cinco millones de hogares estadounidenses recibieron amenazas a través de Facebook el año pasado. Han sido víctimas de virus informáticos, robos de identidad y hostigamientos, informa el estudio que entrevistó a 2.089 hogares en el país.

El grupo de consumidores pide a Facebook reforzar la vigilancia para reducir el número de niños inscritos en la red social.

Fuente: <http://www.nacion.com/2011-05-10/Tecnologia/UltimaHora/Tecnologia2774169.aspx>

Con base en la lectura conteste las siguientes preguntas:

1. ¿Qué problemática social se está denunciando en la noticia?
2. ¿Quiénes están sufriendo esta problemática?
3. ¿Cuáles son las características de los afectados?

### Análisis de la actividad 2

1. Aunque la noticia está referida a una problemática social que afecta a los Estados Unidos, la realidad costarricense no se aleja mucho de lo descrito en esta noticia. La problemática social que se aborda es la amenaza que representan las redes sociales para los niños menores de 13 años y el desentendimiento de sus padres.
2. Los afectados son niños menores de 13 años.
3. La característica principal de los afectados es que no son supervisados por sus padres cuando navegan en las redes sociales.

Es importante que el estudiante pueda extraer información presente en diversos medios de comunicación, fomentando así la comprensión de lectura. El estudio aquí descrito se realizó en los Estados Unidos, en esta noticia solamente se describen algunos resultados; sin embargo, el trasfondo es muy amplio, ya que es casi imposible que pudieran entrevistar a todos los niños menores de trece años que utilizan las redes sociales, lo que hace pensar que utilizaron algunos mecanismos para la recolección y tratamiento de la información de modo que les permitió generalizar los resultados obtenidos sobre la población de los niños estadounidenses menores de 13 años.

Análisis de situaciones como la descrita en este reportaje son importantes insumos para llevar a cabo discusiones en el aula, ya que no solo se ataca la extracción de información sino que puede generar reflexiones entre el profesor y los estudiantes sobre temas de actualidad y que se involucran con el contexto en el que se desarrollan los niños y jóvenes. Además, si se generan estas discusiones el estudiante se dará cuenta que la información y tratamiento estadístico de la misma son una herramienta para entender la sociedad en la que viven.

## Conceptos básicos

### Actividad 3

*Lea cuidadosamente la siguiente información:*

*En una institución de educación secundaria, el director necesita conocer el nivel socioeconómico de los núcleos familiares de sus estudiantes para informar la situación al Ministerio de Educación Pública y conseguir con esto mayor ayuda económica para algunos estudiantes.*

*Para ello decide llevar a cabo una encuesta, donde consulta a los padres o encargados de los estudiantes, entre otras cosas:*

- *El número de personas asalariadas pertenecientes al núcleo familiar.*
- *El salario neto mensual de las personas que conforman el núcleo familiar.*
- *Sexo del jefe del núcleo familiar.*
- *El nivel de escolaridad que tiene el padre, la madre o el encargado del estudiante.*
- *El número de personas en primaria, secundaria o en educación superior que conforman el núcleo familiar.*
- *Si la vivienda donde habitan es alquilada, prestada o propia.*
- *Área en m<sup>2</sup> de construcción de la casa.*
- *El número de personas pertenecientes al núcleo familiar.*

*Después de recolectar los datos, el director los resume, analiza y presenta en un informe a la Dirección Regional de Educación correspondiente.*

*Con base en la situación planteada conteste cada una de las siguientes preguntas:*

1. *¿Cuál es la situación problema a la que tiene que darle respuesta el director de la institución?*

2. ¿A quién o quienes describirá la información recolectada en el estudio?
3. ¿Qué estrategia podría seguir el director para la recolección de la información?
4. ¿Cuáles son las características que se van a analizar?
5. ¿Qué tipo de datos (numéricos o no numéricos) se tendrán al final de la recolección?
6. ¿Qué estrategias puede utilizar el director con los datos obtenidos para llevar a cabo un análisis respectivo y comunicar el mensaje de la mejor manera posible?

### Análisis de la Actividad 3

1. La situación problema que el director de la institución quiere resolver es conocer el nivel socioeconómico de las familias que tienen a sus hijos en dicha institución, para informar al Ministerio de Educación Pública.
2. La información que se recolectará describe el nivel socioeconómico de los núcleos familiares de los estudiantes de la institución.
3. La estrategia de recolección de información puede ser la encuesta, que se aplicará a todas las familias que tienen a sus hijos en la institución.
4. Las características del núcleo familiar sobre las que se va dirigir el análisis son:
  - Número de personas asalariadas pertenecientes al núcleo familiar.
  - El salario neto mensual de las personas que conforman el núcleo familiar.
  - Sexo del jefe del núcleo familiar.
  - El nivel de escolaridad que tiene el padre, la madre o el encargado del estudiante.
  - El número de personas en primaria, secundaria o en educación superior que habitan en la vivienda.
  - La condición de la vivienda donde habitan: alquilada, prestada o propia.
  - Área en m<sup>2</sup> de construcción de la casa.
  - El número de personas que habitan en la vivienda.
5. El listado anterior incluye tanto información numérica como de cualidades o características (no numérica). Esto pues no sólo los aspectos numéricos son relevantes para este estudio, sino que características como la condición de la vivienda, nivel de escolaridad, entre otros, son también de trascendencia para el análisis integral de la situación.
6. Por otro lado, el director de la institución debe prever desde el inicio de su estudio, que los datos que obtendrá tienen diferencias importantes entre los diferentes núcleos familiares, es decir se espera que exista gran *variabilidad* en los datos, por lo que requiere indagar acerca de las estrategias que existen para el resumen, presentación y análisis estadístico de los mismos.

Con la Actividad 3, se han introducido algunos conceptos básicos de Estadística que merecen especial atención, a continuación se presenta cada uno de ellos:

## Unidad elemental o unidad estadística

En primer lugar en todo estudio estadístico se requiere tener muy bien definida la unidad básica que va a proporcionar la información necesaria. En la Actividad 3, la información recolectada proviene del núcleo familiar de los estudiantes de la institución educativa, por lo que él representa la unidad básica de estudio. Esta unidad básica de estudio normalmente recibe el nombre de **unidad elemental o unidad estadística**.

## Características o variables

En una situación o estudio particular, una vez establecida la unidad estadística, solamente algunas de sus particularidades son objeto de estudio, es decir, dentro de todos aquellos aspectos que caracterizan a la unidad elemental únicamente interesan los que se relacionen con el propósito del estudio. En la Actividad 3 se establecieron varias características que fueron listadas: número de personas asalariadas pertenecientes al núcleo familiar, el salario neto mensual de las personas que conforman el núcleo familiar, entre otras. Como se citó anteriormente estas particularidades pueden ser de naturaleza cuantitativa (numérica) y también cualitativa (no numérica), debido a que corresponden a cualidades de las unidades estadísticas. Independientemente de su naturaleza tradicionalmente se les llama **características de las unidades estadísticas**. No obstante, debido a que el valor de una característica varía de una unidad estadística a otra, y esta **variabilidad** se convierte en la principal fuente de análisis estadístico, se acostumbra catalogar a las características con el nombre de *variables*.

## Observación o dato

En la Actividad 3, al momento de aplicar el cuestionario, cada uno de los núcleos familiares proporciona información para cada una de las características o variables en estudio, estos valores se denominan **observaciones** o mejor aún **datos**. Es decir una observación o dato representa el valor numérico o la cualidad o categoría que se obtiene de una unidad estadística para una variable en particular. Por ejemplo, para la variable *número de personas que habitan en la vivienda*, los datos pueden ser 1, 2, entre otros; mientras que para la *condición de la vivienda* las observaciones pueden ser alquilada, prestada o propia. Por lo anterior, para cada característica o variable de interés en un estudio, los datos estadísticos son conjuntos de números o de categorías correspondientes a las observaciones o respuestas obtenidas en el análisis de las diferentes unidades estadísticas incluidas en el estudio.

Por lo anterior, un dato aislado, sino se compara con otros datos, no se considera como una observación estadística.

## Población

Por lo que se ha venido señalando, un estudio de naturaleza estadística involucra una cantidad grande de unidades estadísticas, todas ellas son objeto de estudio. En este sentido, la totalidad de unidades estadísticas recibe el nombre de *población* en estudio. En la Actividad 3, la población está representada por los núcleos familiares de todos los estudiantes de la institución.

En general, una población puede ser finita, infinita o indeterminada. Un ejemplo de un caso de población finita corresponde al problema que se citó en la Actividad 3.

Un ejemplo de población indeterminada, se plantea en el siguiente ejemplo:

*Una empresa que fabrica enlatados de palmito desea analizar si la cantidad de producto por enlatado se ajusta a lo que se incluye en la etiqueta que son 500 ml. El estudio es urgente pues se ha presentado una demanda, donde se afirma que están vendiendo menos producto del que se afirma. Debido a que la empresa no puede parar el proceso de producción para hacer el estudio, como las unidades estadísticas son los enlatados producidos y la variable de interés es la cantidad de palmito por enlatado, resulta imposible determinar la cantidad total de enlatados, entonces se dice que la población es indeterminada.*

Un ejemplo de una población infinita se puede observar en la siguiente situación hipotética:

*Con la intención de determinar el uso que se le da al suelo en el país, se ha decidido seleccionar aleatoriamente diferentes puntos de coordenadas  $(x,y)$  ( $x$ : latitud,  $y$ : longitud) sobre el territorio nacional e identificar el uso que se le da a la tierra en ese punto particular. Aunque el problema es hipotético, puede notarse que la unidad estadística es un punto de coordenadas  $(x,y)$  sobre el territorio nacional, debido a que este punto se elige dentro del continuo de puntos territoriales de Costa Rica, la población de interés es infinita.*

## Muestra

### Actividad 4

Lea cuidadosamente la situación planteada:

*Mario acompaña a su mamá a realizar las compras en el supermercado, cuando pasan por el arroz, su mamá le dice:*

*- Mario, trae una bolsa de arroz que tenga 95% de grano entero.*

*El hijo obedece a su madre y trae la bolsa de arroz y le pregunta:*

*-¿Qué significa esa leyenda de 95% de grano entero?*

*A lo que la mamá le responde*

*- Pues la verdad no sé, a mí lo que me gusta de ese arroz es que se ve muy bonito a la hora de cocinarlo, muy sueltito y entero.*

*Mario llega al otro día al colegio, y le pregunta a su profesor ¿Qué significa que una bolsa de arroz tenga 95% de grano entero?*

*Si usted fuera el profesor de Mario*

- 1. ¿Qué le contestaría?*
- 2. ¿Qué métodos pueden usar en la clase para comprobar si la leyenda en la bolsa es correcta?*

## Análisis de la Actividad 4

1. Que una bolsa de arroz contenga 95% de grano entero significa que, aproximadamente, de cada 100 granos en la bolsa, 95 están enteros.
2. Como no sería viable contar todos los granos de arroz de la bolsa para saber cuántos están completos y cuantos no, se pueden extraer pequeñas muestras del contenido de la bolsa, ya que la distribución de arroz dentro de la bolsa se espera que sea uniforme, lo que indica que el comportamiento en algunas muestras va a poder dar una idea de lo que ocurre con todo el arroz contenido en la bolsa.

Para esto el profesor puede formar subgrupos de cuatro estudiantes para que cada uno tome una muestra de arroz de la bolsa, puede usar como medida el contenido de una tapa de refresco en recipiente de vidrio de 375 ml (comúnmente llamada como chapa), todos los estudiantes de cada subgrupo realiza un conteo de granos enteros y partidos. Luego se hace la proporción por subgrupo y luego de todo el grupo para poder ir creándose una idea acerca de si el mensaje en la bolsa es viable o no.

El procedimiento que se hizo para la obtención de las muestras no es aleatorio por lo que el experimento no se puede generalizar a toda la población; sin embargo, el estudiante puede darse una idea de lo que sucede con el contenido de la bolsa.

En la Actividad 4 se utiliza el concepto de **muestra**, la cual es una parte de la población. Se extrae para ser analizada y calcular algunas medidas estadísticas con objetivos claramente establecidos antes del estudio que se desee realizar.

Los elementos de una muestra pueden obtenerse de la población siguiendo un proceso de aleatoriedad donde las unidades estadísticas se seleccionan tomando en cuenta estrategias probabilísticas o mediante un proceso no aleatorio donde se eligen por conveniencia o facilidad.

El procedimiento de selección de los elementos que pertenecen a la muestra depende del tipo de información que se quiere obtener de la muestra, las diferencias o similitudes que tengan las características de la población y las facilidades que tenga quien investiga, como tiempo, recursos económicos, entre otros. Es importante rescatar que si se quieren hacer inferencias sobre la población utilizando una muestra como referencia, esta última debe seleccionarse de manera aleatoria.

### **Ejemplos**

Cuando una persona quiere hacerse un diagnóstico de salud, normalmente el médico inicia enviándole un examen de sangre, de heces y orina; de este modo el paciente proporciona una pequeña muestra de cada una de estas sustancias, por medio de las cuales mediante un estudio detallado es posible identificar su estado de salud en relación con una serie de características relacionadas con un adecuado funcionamiento del organismo. A pesar que las muestras aportadas son pequeñas, con ellas se pueden realizar inferencias importantes sobre el estado de diferentes órganos del cuerpo. Este es un ejemplo de la forma en que el muestreo ayuda para la toma de decisiones respecto al comportamiento de una serie de variables. La pequeña cantidad de producto que se utiliza se debe a que las características de cada uno de ellos presentan poca variación. Por ejemplo, los mililitros que obtienen de sangre dan la información de toda la sangre que circula por el cuerpo, ya que ésta es muy homogénea.

Por otra parte, cuando se realizan las elecciones presidenciales en el país, hay instituciones encargadas de realizar encuestas que determinan las preferencias de los votantes conforme se vaya acercando la fecha de la elección. Por razones de costo y

tiempo, dichas encuestas no se pueden realizar al total de la población, por eso las instituciones responsables de estos estudios realizan encuestas a muestras relativamente pequeñas (normalmente entre 1000 y 2000 personas, a pesar que hay cerca de tres millones de electores).

También se presentan situaciones en las cuales el muestreo es absolutamente necesario, pues la obtención de los datos estadísticos implica la destrucción de la unidad estadística, un ejemplo de esto lo representa el análisis de los enlatados de palmito que se citó anteriormente, para analizar el contenido de los enlatados el producto debe ser abierto y pesado. Muchos otros estudios presentan el mismo problema, por lo que el proceso de muestreo adquiere una relevancia fundamental.

Es importante hacer notar que entre más variabilidad posean las características de las unidades estadísticas más grande debe ser la muestra que se selecciona. De hecho, la variabilidad es más importante que el tamaño de la población al momento de seleccionar una muestra para un estudio particular.

## Variabilidad

### Actividad 5

*Lea atentamente la siguiente situación y conteste la pregunta planteada:*

*El profesor de Matemática pide a un grupo de cuatro estudiantes de la sección 8-1 de un colegio particular que caractericen los objetos personales que pueden observarse del resto de los compañeros, entre ellos: color de camisas y blusas, color de pantalón o enagua, color de las mochilas, largo del cabello y sexo.*

*Los estudiantes deben realizar la recolección de datos por medio de la observación, resumirlos, analizarlos y presentarlos al profesor.*

*Los estudiantes hacen una tabla con el nombre de sus compañeros y las características observadas como se presenta en el cuadro adjunto.*

**Cuadro 1.** Caracterización de los estudiantes de la sección 8-1

Nombre	Sexo	Color de camisas y blusas	Color de pantalón o enagua	Color de las mochilas	Largo del cabello
María José	M	Celeste	Azul	Rosada	Largo
Arturo	H	Celeste	Azul	Azul	Corto
Antonio	H	Celeste	Azul	Morado	Semi corto
Sofía	M	Celeste	Azul	Verde	Corto
...	...	...	...	...	...

*Luego piden orientación al profesor, entre las preguntas que hacen están:*

*-Como no podíamos medir nada solo observar, en el largo del cabello lo que hicimos fue clasificarlo como:*

*Corto: si está pegadito a la cabeza o arriba de los hombros*

*Semi corto: si está por los hombros*

*Largo: si está por debajo de los hombros*

*¿Verdad que se podía hacer así?*

*A lo que el profesor responde:*

*-Claro que sí, esa era la idea que ustedes encontrarán la estrategia para clasificar los datos que obtuvieron.*

*Los estudiantes continúan*

*-Con el color de las camisas y blusas, no tenemos ninguna duda ya que es muy fácil la clasificación TODAS son de color celeste, porque así es el uniforme de nuestra institución. Ocurre algo similar con el color del pantalón o enagua ya que TODOS son azules.*

*Ahora cómo hacemos para contestar a la pregunta ¿De qué color es la mochila de los estudiantes de la sección 8-1? Ya no podemos dar una respuesta directa. Y ¿Cuáles son las diferencias de esas variables que inducen a que unas sean más fáciles de analizar que otras?*

*Si fueras el profesor de esos estudiantes ¿Cuáles serían sus respuestas?*

## **Análisis de la Actividad 5**

Es importante rescatar que el profesor no debe dar de una vez la respuesta a este tipo de interrogantes sino que debe dirigir una discusión con los estudiantes para que ellos mismos logren encontrar la respuesta.

El docente puede hacer las siguientes preguntas:

*¿Qué propiedad se puede observar en los datos que arrojó la variable “color de camisa o blusa”?*

*Los estudiantes pueden decir*

*-pues que todas son iguales.*

*¿Qué propiedad se puede observar en los datos que arrojó la variable “color de la mochila”?*

*Los estudiantes pueden responder*

*-pues que no todos son iguales, hay diversidad en las respuestas.*

*El profesor puede concluir:*

*-¡pues ahí está el punto! En la primera de las características No hay variabilidad todas son iguales; sin embargo, en la segunda las repuestas son diversas por lo que se dice que hay variabilidad en los datos.*

*El profesor continúa*

*-Ya saben que el color de la mochila no se puede generalizar a todo el grupo ya que hay variabilidad, entonces ¿qué pueden hacer? ¿Saben cuántos colores distintos hay en las mochilas? Para cada color ¿pueden determinar cuántas mochilas hay?*

*Con este tipo de preguntas los estudiantes se irán dando una idea de cómo pueden hacer la representación de datos con variabilidad, ya sea en cuadros o gráficas estadísticas<sup>1</sup>.*

Estas situaciones tienen particular relevancia en los estudios estadísticos, el análisis de la variabilidad en los datos es el principal fundamento de la disciplina; las técnicas estadísticas que se han desarrollado tienen como principal fundamento ayudar a describir, proyectar y analizar estos patrones de variabilidad de los datos. Por ejemplo, para efectos de la Actividad 5, para el color de la camisa o blusa, basta con indicar que todos los estudiantes utilizan el mismo color para cerrar el estudio; mientras que para el análisis del color de la mochila se requiere establecer técnicas de resumen de información, cuadros o gráficas para evidenciar el patrón de variabilidad y para poder llevar a cabo una reflexión sobre dicho patrón. Por esta razón, en un estudio estadístico el concepto de *variabilidad* juega un papel preponderante y debe ser tomado en cuenta en todo momento.

\*\*\*

## Características o variables

Anteriormente se definió lo que era una característica o variable, a continuación se retoma el tópico y se amplía.

### Actividad 6

*Retomando la actividad donde el director requiere recopilar información del nivel socioeconómico de las familias que tienen a sus hijos en la institución que él dirige. Sabemos que en dicha actividad se estableció que las características del estudio son:*

- *Número de personas asalariadas pertenecientes al núcleo familiar.*
- *El salario neto mensual de las personas que conforman el núcleo familiar.*
- *El nivel de escolaridad que tiene el padre, la madre o el encargado del estudiante.*
- *El número de personas en primaria, secundaria o en educación superior que habitan en la vivienda.*
- *La condición de la vivienda donde habitan: alquilada, prestada o propia.*
- *Área en m<sup>2</sup> de construcción de la vivienda.*
- *El número de personas que viven en la vivienda.*

1. *Para cada una de estas variables determine si los datos obtenidos son numéricas o cualidades.*
2. *En relación con las variables numéricas indique si los datos correspondientes se obtienen por conteo o por medición.*
3. *En relación con las variables cualitativas indique si los datos correspondientes pueden ser ordenados siguiendo algún patrón u orden natural entre ellos.*

<sup>1</sup> En este material didáctico, y en la propuesta de los nuevos programas de estudio, considere como sinónimos las palabras tabla y cuadro también gráfica y gráfico estadístico.

## Análisis de la Actividad 6

En el cuadro 2 se encuentra la clasificación de las variables según la condición de numéricas y de cualidad.

**Cuadro 2.** Clasificación de las variables según el tipo de datos que se obtuvieron

<b>Variables o características</b>	<b>Tipo de datos</b>
Número de personas asalariadas pertenecientes al núcleo familiar	Numérica
Salario neto mensual de las personas que conforman el núcleo familiar	Cantidad
Número de personas en primaria, secundaria o en educación superior que habitan en la vivienda	Cantidad
Condición de la vivienda donde habitan: alquilada, prestada o propia	Cualidad
Área en m <sup>2</sup> de construcción de la vivienda	Cantidad
Número de personas que viven en la vivienda.	Cantidad
Nivel de escolaridad que tiene el padre, la madre o el encargado del estudiante	Cualidad

En el cuadro 3 se encuentra la clasificación de las variables numéricas.

**Cuadro 3.** Clasificación de las variables numéricas

<b>Variables numéricas</b>	<b>Forma de obtener los datos</b>
Número de personas asalariadas pertenecientes al núcleo familiar	Conteo
Número de personas en primaria, secundaria o en educación superior que habitan en la vivienda	Conteo
Número de personas que viven en la vivienda.	Conteo
Área en m <sup>2</sup> de construcción de la vivienda	Medición
Salario neto mensual de las personas que conforman el núcleo familiar	Medición

En el cuadro 4 se encuentra la clasificación de las variables cualitativas.

**Cuadro 4.** Clasificación de las variables cualitativas

Características o cualidades	Propiedades de las variables
Sexo del jefe del núcleo familiar	No hay orden natural
Condición de si la vivienda donde habitan es alquilada, prestada o propia	No hay orden natural
Nivel de escolaridad que tiene el padre, la madre o el encargado del estudiante	Si hay un orden natural

Dentro de un análisis estadístico si los valores posibles de la variable pueden ser expresados numéricamente entonces a estas variables se le llaman **variables cuantitativas**. No obstante, dentro de este tipo de características de las unidades estadísticas también surge una subclasificación que tiene relevancia técnica, por ejemplo hay diferencias en la forma de obtener los datos para las variables número de personas asalariadas en el núcleo familiar y área en m<sup>2</sup> de construcción de la vivienda, en el primer caso los datos se obtienen por conteo, por lo que este tipo de variables toman generalmente únicamente números enteros y se les denomina **variables discretas**. Para el segundo caso, el área de construcción debe ser obtenido por medición, en este caso se utiliza un instrumento para medición de área y la unidad es el metro cuadrado; estas variables teóricamente podrían tomar cualquier valor real en un intervalo determinado, este tipo de variables se denominan **variables continuas**, y normalmente se requiere un instrumento de medición para la obtención de los datos

## Variables cualitativas nominales y ordinales

En relación con las características no numéricas de las unidades estadísticas, se denominan **variables cualitativas**, también es posible realizar una subclasificación en relación con el tipo de dato que generan. Como puede notarse en la variable sexo del jefe del núcleo familiar, los datos pueden variar en dos categorías: hombre o mujer; pero estas dos categorías pueden ser colocadas en cualquier orden debido a que no existe prioridad clara entre ellas, a estas variables se les denomina **variables cualitativas nominales**. Por otro lado, cuando se analiza la variable nivel de escolaridad de la persona encargada del estudiante, se tiene claro que las categorías de escolaridad aunque son cualidades, pueden ser ordenadas naturalmente de menos escolaridad a más escolaridad o viceversa. En los casos en que se presenta esta situación se dice que la característica es una **variable cualitativa ordinal**.

Algunos ejemplos de variables cualitativas nominales son: color de los ojos, religión, preferencia política, entre otros. Por su parte, algunos ejemplos de variables cualitativas ordinales son: tallas de las camisas (S, M, L), condición socioeconómica (baja, media, alta), entre otras.

**Nota:** A las variables cualitativas se les acostumbra llamar variables categóricas, pues los datos que generan pertenecen a diferentes categorías. Entonces cuando una variable cuantitativa se agrupa en clases puede ser analizada como una variable categórica.

**Ejemplo**

Se desea estudiar, entre otras variables, el peso y la estatura de los estudiantes de un colegio, con el propósito de evaluar su condición de salud. Una vez recolectada la información, con los datos se calculó el Índice de Masa Corporal (IMC) que viene dado por la fórmula:

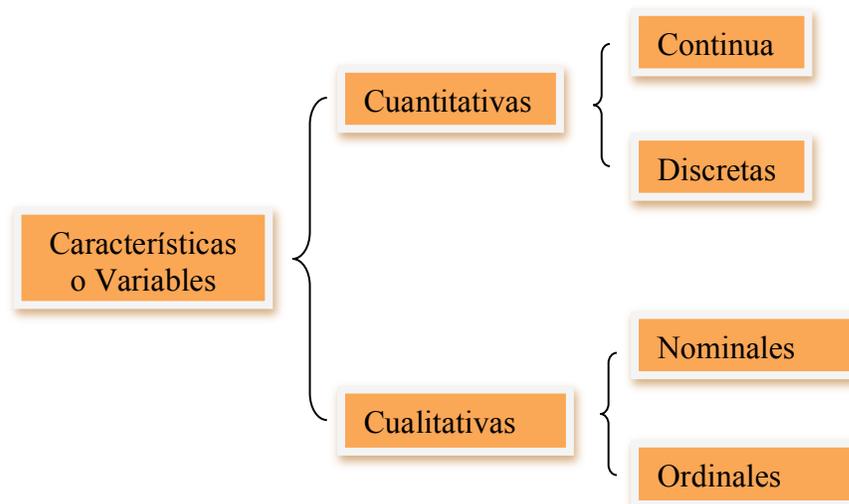
$$IMC = \frac{Peso(kg)}{Estatura(m)^2}$$

Una vez determinados los valores numéricos de este índice, tomando en cuenta la edad y el sexo de los estudiantes, se realizó la siguiente clasificación:

- Peso insuficiente
- Peso normal
- Sobrepeso
- Preobesidad
- Obesidad I
- Obesidad II
- Obesidad extrema

En este caso se puede notar que la variable IMC que originalmente es numérica se reclasifica en una variable categórica; pero no pierde su sentido pues la condición de que un estudiante pertenezca a una categoría está en función del valor numérico de su IMC.

A manera de resumen, el concepto de característica o variable es lo que deseamos medir o evaluar de la unidad estadística, se puede realizar una clasificación de las mismas, como se presenta en el siguiente esquema:



## Actividad 7

Determine para cada variable contextualizada su clasificación, a saber: cualitativa (nominal u ordinal) y cuantitativa (continua y discreta).

<b>Variable contextualizada</b>	<b>Clasificación de la variable</b>
<i>Estaturas de los integrantes del equipo de baloncesto de un colegio.</i>	
<i>Número de hermanos de cada uno de los estudiantes de séptimo grado de un colegio en particular.</i>	
<i>Deporte preferido de los estudiantes de undécimo año.</i>	
<i>Color de zapatos de los asistentes a una fiesta.</i>	
<i>Nivel de escolaridad de los padres de familia de los estudiantes de un grupo guía del docente.</i>	
<i>Longitud que hay entre el dedo anular y el codo de los alumnos del aula.</i>	
<i>Cantidad de padres de familia asistentes a la primera reunión del año lectivo por nivel, en una institución en particular.</i>	
<i>Temperatura de una persona o su estado febril utilizando la clasificación Subfebril o febrícula: Menos de 37,5°C; fiebre ligera: Menos de 38°C; fiebre moderada: 38 – 39°C; fiebre alta: 40°C y hiperpirexia: 41°C.</i>	

## Análisis de la Actividad 7

<b>Variable contextualizada</b>	<b>Clasificación de la variable</b>
<i>Estaturas de los integrantes del equipo de baloncesto de un colegio.</i>	<i>Cuantitativa continua</i>
<i>Número de hermanos de cada uno de los estudiantes de séptimo grado de un colegio en particular.</i>	<i>Cuantitativa discreta</i>

Deporte preferido de los estudiantes de undécimo año	<i>Cualitativa nominal</i>
Color de zapatos de los asistentes a una fiesta.	<i>Cualitativa nominal</i>
Nivel de escolaridad de los padres de familia de los estudiantes de un grupo guía del docente.	<i>Cualitativa ordinal</i>
Longitud que hay entre el dedo anular y el codo de alumnos del aula.	<i>Cuantitativa continua</i>
Cantidad de padres de familia asistentes a la primera reunión del año lectivo por nivel, en una institución en particular.	<i>Cuantitativa discreta</i>
Temperatura de una persona o su estado febril utilizando la clasificación Subfebril o febrícula: Menos de 37,5°C; fiebre ligera: Menos de 38°C; fiebre moderada: 38 – 39°C; fiebre alta: 40°C y hiperpirexia: 41°C.	<i>Cualitativa ordinal</i>

Hasta este momento se han abarcado los conceptos básicos de la Estadística. Dicha área se subdivide en Estadística descriptiva y Estadística inferencial, a continuación se hace una descripción más detallada de estas.

**Estadística descriptiva:** es una parte de la Estadística que se dedica a organizar, representar, resumir y analizar conjuntos de datos, de modo que se pueda describir en forma precisa las características de interés en función del contexto al que pertenecen. Este análisis es muy elemental, pero fundamental en cualquier estudio que requiera de datos para la toma de decisiones.

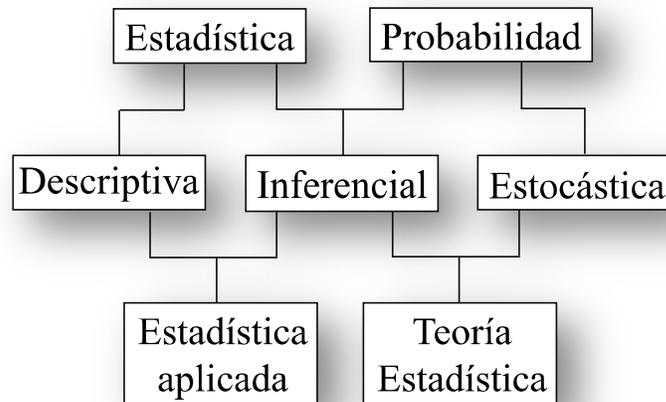
En esta etapa la confección de cuadros y gráficas, el cálculo de medidas de posición como de variabilidad, así como el cálculo de porcentajes son algunas de las técnicas comúnmente utilizadas, para favorecer la interpretación, el análisis y la argumentación, que se genera de los datos. Cabe rescatar que los distintos tipos de representación de datos son un apoyo para la interpretación y análisis de los mismos, su construcción no debe ser el fin del análisis estadístico.

**Estadística inferencial:** debido al propósito básico de la Estadística se fundamenta en identificar los patrones de variabilidad de los datos para favorecer su interpretación, y con ello poder apoyar el análisis de situaciones particulares relacionadas con fenómenos vinculados con las diferentes áreas del conocimiento. En este sentido, la información proporcionada por los análisis descriptivos en una muestra aleatoria de datos, pueden ser utilizados para extraer conclusiones sobre un conjunto mayor (población de donde se extrajo esa muestra). Esta capacidad para poder inferir sobre la población con base en los resultados de una muestra al azar está basada en supuestos probabilísticos relacionados con los patrones de distribución de los datos.

El conjunto de técnicas y procedimientos por medio de los cuales se vinculan los resultados de la Estadística descriptiva con la aplicación de modelos de probabilidad,

para inferir o generalizar esos resultados es denominado como Inferencia estadística o Estadística inferencial.

Finalmente, se presenta un esquema relacionado con la Estadística, su relación con la Probabilidad y demás áreas afines.



La Estadística y la Probabilidad a través de la historia surgen como dos ramas completamente separadas; sin embargo, años después se descubre que si se unen se pueden crear nuevas áreas de conocimiento. La Estadística como tal se subdivide en descriptiva e inferencial (estadística aplicada). La Estadística inferencial es el vínculo entre la Estadística y la Probabilidad, aquí se trabaja con muestras obtenidas bajo procesos probabilísticos y gracias a esta obtención se pueden hacer inferencias sobre las poblaciones en estudio.

La Probabilidad por si sola contempla procesos estocásticos y si consideramos la mezcla que existe entre inferencia y estos procesos obtenemos la teoría estadística.

## III. Recolección de información

### Introducción

En la sección anterior quedó en evidencia que en los estudios estadísticos los datos constituyen el objeto básico; pero además, se indicó que a partir de observaciones aisladas no es posible efectuar esos análisis. Por esta razón, una mención especial requieren los procesos de búsqueda y recolección de datos, debido a que la calidad de los estudios estadísticos va a depender, en primera instancia, de la eficacia de la información y de la forma en que se ha recolectado. Debido a lo anterior, este tema va a ser analizado en la presente sección, junto con los principales aspectos vinculados con las estrategias didácticas para la enseñanza de este tema.

### Actividad 1

*Analice la siguiente situación para introducir a las y los estudiantes en la búsqueda y recolección de información.*

*Suponga que un docente de III Ciclo está muy interesado en conocer los factores que intervienen en el bajo rendimiento en matemáticas de sus estudiantes. El docente sabe que para llevar a cabo ese estudio requiere de datos de calidad y aplicables a su contexto. Una primera interrogante que le podría surgir es ¿qué variables intervienen en el bajo rendimiento? Y para ello necesitará realizar una adecuada búsqueda bibliográfica.*

*¿Qué opciones de recolección de información se le podrían recomendar al profesor?*

### Análisis de la Actividad 1

Partiendo del hecho que el docente previamente planteó el problema a investigar y delimitó los objetivos de su estudio, se debe concentrar la atención en el proceso de búsqueda y recolección de los datos. Una primera sugerencia consiste en realizar una búsqueda bibliográfica sobre el tema; para ello podría recurrir a diversas publicaciones en revistas educativas físicas o electrónicas, tesis, artículos publicados en Internet, estadísticas sobre la temática recopiladas por alguna institución como por ejemplo, el Estado de la Nación, el Departamento de Análisis Estadístico del MEP o el Departamento de Estudios e Investigación Educativa de la Dirección de Planificación Institucional del MEP. En general, la primera etapa debería ser un proceso de búsqueda bibliográfica en diversas investigaciones que abarquen el tema de interés.

En una segunda etapa, una vez que se haya identificado la información existente sobre el fenómeno en estudio, se requiere establecer alguna estrategia para recolectar aquella información sobre la cual no existen datos. En este caso, el docente debería recurrir al uso de algunas técnicas para la recolección de la información que no está a su disposición. Por ejemplo, se podría construir un pequeño cuestionario y administrarlo a sus estudiantes, también podría observar a ciertos estudiantes, durante varias de sus lecciones, para identificar factores que intervienen en el bajo rendimiento en Matemática; otra opción podría ser entrevistar de manera directa a algunos de sus alumnos. También podría simular un pequeño experimento definiendo dos grupos de

estudiantes, uno actuaría como grupo control mientras que el otro estaría expuesto a diversos estímulos tales como una enseñanza más individualizada, trabajo paralelo con los padres de familia que denote un verdadero interés hacia la situación escolar de sus hijos, uso de recursos didácticos innovadores, entre otros.

De igual manera el docente puede plantear diferentes situaciones en el contexto de los estudiantes, que los hagan reflexionar sobre la forma de recolectar información que está disponible o incluso que hay que recolectar de alguna manera porque no se encuentra a disposición, para este caso se podría optar por técnicas como la interrogación, la observación y la experimentación.

## **Generalidades acerca de la recolección de información**

La Estadística es una herramienta utilitaria en muchos campos del saber, su uso se focaliza en dos fines: primeramente para apoyar la investigación científica dirigida a generar conocimiento sobre un fenómeno particular; en segundo lugar para la toma de decisiones en diferentes campos. Independientemente del fin, tal como se ha venido mencionando, los datos constituyen el objeto de estudio; por lo que las estrategias de recolección de éstos deben ser cuidadosamente seleccionadas y seguir criterios científicos.

En la práctica, pueden presentarse dos situaciones relacionadas con la recolección de información:

- a) Los datos requeridos para un estudio existen y están disponibles en diferentes fuentes
- b) Los datos requeridos no están disponibles y deben ser recabados.

En el primer caso, se habla de información existente, cuando un investigador o institución la recogió para estudiar el mismo o un problema similar al que se está proponiendo o porque los datos son recogidos y publicados periódicamente por alguna institución. Por otro lado, cuando los datos no están disponibles se debe definir y ejecutar una estrategia para obtenerlos.

## Información existente

Si la información requerida para un estudio está a disposición se pueden distinguir dos tipos de fuentes a las cuales recurrir: las **primarias** y las **secundarias**.

### Actividad 2

*Analice la siguiente situación.*

*El Instituto Nacional de Estadísticas y Censos (INEC) es la institución encargada de realizar el Censo de población y vivienda; así como encuestas que sistemáticamente se aplican a diferentes sectores de la población; por ejemplo, hasta el año 2009 se venía realizando la Encuesta de Hogares de Propósitos Múltiples que en el 2010 se sustituyó por un nuevo ciclo de encuestas de hogares denominado Encuesta Nacional de Hogares. Este instrumento recopila datos de una serie de variables de interés nacional. El INEC organiza y resume la información que finalmente se presenta en cuadros estadísticos desagregados por sexo, grupos de edad, zona urbana, rural y regiones de planificación.*

*La Comisión Económica para América Latina y el Caribe (CEPAL) realizó un estudio sobre Inmigración en Costa Rica: características sociales y laborales, integración y políticas públicas. Ese estudio abordó numerosas dimensiones de ese fenómeno social en el país, desde la caracterización demográfica y socioeconómica de los migrantes hasta el marco jurídico que lo norma. La CEPAL tomó los datos de dos fuentes principales: el Censo de población y vivienda del 2000 y la Encuesta de Hogares de Propósitos Múltiples del 2007.*

*De acuerdo con el texto anterior, responda las siguientes preguntas:*

- 1. Si usted requiere información para un estudio y utiliza los datos censales que publica el INEC en la página Web: [www.inec.go.cr](http://www.inec.go.cr). ¿Está usted consultando una fuente primaria o una fuente secundaria?*
- 2. En la misma situación de la pregunta anterior, si ahora consulta el estudio que realizó la CEPAL sobre inmigración en Costa Rica ¿Está usted consultando una fuente primaria o una fuente secundaria?*

### Análisis de la Actividad 2

- 1. Debido a que el INEC recaba directamente la información utilizando procedimientos que ellos mismos establecen y ponen en práctica, entonces este hecho convierte al INEC en una fuente de información primaria para cualquier consulta sobre el censo o cualquier otra encuesta que ellos realizan sistemáticamente.*
- 2. Por su parte, debido a que la CEPAL utilizó información de fuentes primarias para su estudio sobre inmigración, entonces al emplear los resultados de este estudio se está haciendo uso de una fuente secundaria, pues la CEPAL no*

recabó información por ellos mismos, sino que emplearon información que otras instituciones recabaron.

## Fuentes primarias y secundarias

**Una fuente primaria** corresponde a aquella oficina, persona o agencia que publica o proporciona los datos que ella misma ha producido. Se trata del productor original de la información; por ejemplo, el INEC (censo, estadísticas de población, Encuesta de Hogares, vivienda, comercio, estadísticas vitales, etc.), el Banco Central de Costa Rica (estadísticas de balanzas de pagos, monetarias, bancarias, etc.).

Una fuente no es necesariamente una publicación. Lo que se requiere es que los datos hayan sido recogidos, tabulados y que existan en forma tal que permita su suministro (por parte de quien los recolectó) y su utilización por parte del interesado.

**Una fuente secundaria** se refiere a aquella oficina, persona o agencia que hace publicaciones o trabajos que resumen o presentan información originalmente recopilada por otros que son los productores del dato; por ejemplo, cuando la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) publica datos sobre Costa Rica o los Anuarios publicados por la ONU.

Cuando se realizan investigaciones se recomienda el uso de fuentes primarias por varias razones: generalmente incluyen mayor detalle, menor posibilidad de errores de transcripción, ofrecen definiciones de términos y conceptos empleados, normalmente incluyen un ejemplar del cuestionario y se realiza una descripción de los procedimientos usados, explicación de la forma de capacitación del personal entrevistador, forma en que fueron procesados los datos, entre otras.

En todo estudio debería indicarse la fuente (si corresponde) de la cual se tomaron los datos con el mayor detalle posible, a menos que se obtuviera bajo la condición de no revelar su origen; en este caso, se debería señalar "**Fuente: Confidencial**".

Independientemente de la forma en que se han obtenido los datos, es fundamental llevar a cabo un proceso de evaluación de la calidad de la información que se va a utilizar. Evaluar las diversas fuentes permite conocer la confianza que merecen en cuanto a calidad, comparabilidad, pertinencia, exactitud, limitaciones y aplicabilidad para el estudio o investigación que se esté realizando. Seguidamente se plantea una situación que un docente podría utilizar para analizar el tema de información existente y no existente

### Actividad 3

Analice la siguiente situación.

*Un docente de III o IV Ciclo asigna un pequeño proyecto a sus estudiantes para ser trabajado en subgrupos de cuatro integrantes pero que además, incluye la participación de los padres de familia. Los jóvenes tienen la libertad de escoger con quién quiere trabajar, forman los subgrupos y seleccionan un coordinador. El profesor confecciona una guía para el proyecto y la entrega a cada coordinador de subgrupo. El proyecto es un trabajo extracurricular que debe ser expuesto ante los demás compañeros en una fecha que fija el docente.*

#### *Guía del trabajo*

- 1. Piensen, en conjunto con sus padres, en una problemática social actual de nuestro país que le preocupa a los jóvenes.*
- 2. Definan un tema para su proyecto que abarque la problemática seleccionada y muéstrela a su profesor para el visto bueno (la idea es que cada subgrupo trabaje con una temática distinta).*
- 3. Una vez que el docente ha dado el visto bueno para el proyecto, establezca al menos tres objetivos que al subgrupo le interese conocer sobre la temática planteada.*
- 4. Defina qué tipo de datos necesitan para lograr sus objetivos.*
- 5. Consulte a sus padres y profesores ¿dónde pueden obtener información sobre la temática de interés?*
- 6. Con la ayuda de sus padres y profesores, busquen y recolecten la información y datos requeridos para su proyecto.*
- 7. Resuman, clasifiquen, organicen, analicen e interpreten la información recopilada mediante cuadros y gráficas. Muéstrela el avance de su trabajo al profesor.*
- 8. Una vez que el docente le ha dado recomendaciones, preparen una pequeña exposición oral para presentar y defender su proyecto ante los demás compañeros.*

## Análisis de la Actividad 3

Las temáticas que escojan los estudiantes pueden ser muy variadas. Podrían seleccionar problemáticas sociales tales como trabajo infantil, inseguridad ciudadana, violencia familiar, indicadores de pobreza, economía nacional, entre otros. Sin embargo, es posible que los datos que se requieran estén disponibles en alguna institución, o en alguna publicación que se haya realizado. Una vez que el docente conoce las temáticas seleccionadas puede recomendarles un listado de fuentes de recolección de información al que pueden acudir. Por ejemplo, la administración del colegio o al Departamento de Orientación, Departamento de Análisis Estadístico del MEP, al Departamento de Estudios e Investigación Educativa de la Dirección de Planificación Institucional, al INEC, al Banco Central de Costa Rica, las Municipalidades, el Tribunal Supremo de Elecciones, el Ministerio de Trabajo, el Instituto Mixto de Ayuda Social, la Caja Costarricense del Seguro Social, el Ministerio de Salud, entre otros. Es importante resaltar que mucha información de las fuentes mencionadas se encuentra en las páginas de Internet respectivas; con ello los padres de familia y los docentes pueden ayudar y orientar a los jóvenes en la búsqueda de los datos que necesitan. Posteriormente, el profesor podría guiar a las y los estudiantes en la búsqueda de estrategias que le permitan realizar un análisis estadístico de los datos recolectados.

Es importante señalar que el docente debe delimitar la actividad a las posibilidades de datos existentes, de manera que los estudiantes tengan la posibilidad de encontrar la información sin grandes complicaciones, ni grandes traslados.

\*\*\*

La situación anterior ilustra casos donde la información requerida está disponible porque ha sido recogida en otros momentos. No obstante, también se podría generar situaciones en donde el estudiante tenga que utilizar estrategias para recolectar información que no está disponible. Una recomendación de ese tipo se ilustra en la actividad siguiente.

## Información no disponible

### Actividad 4

*Analice las siguientes situaciones que se desarrollan en un colegio hipotético, en el cual hay dos grupos de séptimo año.*

*a) El docente de Matemática de séptimo año está interesado en identificar aquellos estudiantes que tienen problemas de socialización que afectan el trabajo en subgrupos, para ello decide plantear varias situaciones didácticas para que sean resueltas por medio del trabajo grupal, y lleva a cabo un guía de observación para identificar el comportamiento de cada uno durante el proceso. Con esta estrategia puede identificar a los estudiantes que tienen más problemas para socializar.*

*b) Para complementar el estudio realizado en el punto a) el docente decide reunirse con los padres de familia o encargados de los jóvenes que presentaron problemas y realizar algunas preguntas preestablecidas para obtener más información al respecto. De este modo, en conjunto con el Departamento de Orientación, se pueden buscar medidas que les ayuden a esos estudiantes*

c) Debido a que un importante grupo de estudiantes de séptimo año regularmente llega tarde cuando tienen clases por la mañana, el profesor de Matemática decide realizar un estudio para identificar los medios de transporte que utilizan para llegar a la institución. Debido a que el colegio se encuentra frente a una carretera principal cerca de paradas de autobuses, el profesor decide llegar a las 6:00 am en los días en que los grupos de séptimo tienen horario por la mañana para identificar el medio de transporte que utiliza cada uno de los 67 estudiantes de ese nivel. De este modo, permanece en la entrada de la institución hasta las 7:30 am observando y anotando si el estudiante llega caminando, en vehículo particular, en autobús público o en transporte privado. Sin embargo, la orientadora del nivel le sugiere que, para facilitar su trabajo, durante las lecciones de Matemática, dedique un espacio para consultar a cada estudiante por el medio de transporte que regularmente utilizan para trasladarse a la institución.

Con base en la situación planteada discuta:

1. ¿Qué técnicas de recolección de información se ponen en práctica en cada uno de los puntos anteriores?
2. ¿Si usted fuera ese profesor, cuál técnica emplearía en la tercera situación y por qué?
3. ¿Qué ventajas y desventajas observa usted a cada uno de esas técnicas identificadas en las tres situaciones?

#### Análisis de la Actividad 4

1. En la situación a) el profesor recurre a la *observación* para identificar datos relacionados con los patrones de conducta de los estudiantes cuando se trabaja en subgrupos. Esta es una técnica muy utilizada en investigaciones sociales y sobre todo en investigación educativa. También, puede ser empleada en procesos de recolección de información en los trabajos de aula durante las lecciones de Estadística y Probabilidades.
2. En la situación b) se recurre a la *interrogación* a los padres o encargados de los estudiantes con la intención de recolectar datos que le ayuden a complementar las observaciones realizadas. Para ello elaboró una serie de preguntas que aplicó a estas personas, con el propósito de obtener información de naturaleza similar y poder sistematizarla.
3. En la situación c), primeramente se considera oportuno utilizar la *observación* para recolectar esta información, pero después se sugiere que se utilice la *interrogación*. Aquí se evidencia como pueden ser utilizadas diferentes técnicas para recolectar un mismo tipo de datos.
4. Aunque las dos técnicas generan los mismos datos, no necesariamente son igualmente efectivos ni producen observaciones con la misma calidad. Como puede notarse el uso de la observación en la situación c) genera un gran esfuerzo de trabajo, que implica que el docente invierta una hora y treinta minutos cada uno de los días en que va a aplicar la estrategia; pero además, una alta concentración en la actividad, pues no se le puede pasar ningún estudiante sin que registre el medio

de transporte empleado. Por su parte, la interrogación, en este caso, es de más fácil aplicación, pues basta con plantear, en forma directa o mediante alguna estrategia indirecta, una o más preguntas destinadas a recabar esta información.

5. En cuanto a la observación sus ventajas son:

- a) Se recaba la información directamente, por lo que la calidad de los datos está garantizada, siempre que el observador esté debidamente preparado para llevar a cabo este proceso.
- b) La observación permite desarrollar una descripción de los hechos, al mismo tiempo analizar la situación.
- c) Esta técnica es de vital importancia cuando se llevan a cabo experimentos de cualquier tipo.

Las principales desventajas de la observación:

- a) Se requiere una adecuada preparación de las personas que vayan a realizar las observaciones de la forma más natural posible.
- b) Muchas veces, la persona observada puede sentirse intimidada y no actuar naturalmente.
- c) Se requiere definir un instrumento para recabar los datos de cada observación, lo cual no siempre es una tarea fácil.

En cuanto a las ventajas de la interrogación:

- a) Permite unificar información para las distintas unidades estadísticas, pues a todas ellas se les realizan las mismas preguntas.
- b) Debido a que las preguntas son estándar, la preparación del entrevistador no requiere ser tan detallada como en el caso de la observación.
- c) Es adaptable a diferentes situaciones.

Pero la interrogación también tiene desventajas, tales como:

- a) Es difícil obtener información confidencial pues puede ocurrir que el entrevistado se intimide con la pregunta.
- b) Se limita la posibilidad de profundizar sobre un tema, pues la pregunta está dirigida para recolectar información muy específica.

## Actividad 5

- a) *Normalmente se cree que si en un sorteo sale favorecido un número particular, entonces su probabilidad de ocurrencia disminuye para el siguiente sorteo. O en su defecto, si un número no ha salido favorecido en mucho tiempo, se cree que tiene más probabilidad que otros números de salir. Esta creencia se conoce como falacia del jugador, pero también es aplicable a la vida real y afecta el actuar de las personas. Con el propósito de eliminar esta creencia equivocada, un educador pide a los jóvenes que realicen la siguiente experiencia.*

*Tomen un dado numerado del uno al seis, láncelo una primera vez y observen el resultado y luego láncelo una segunda vez y vuelvan a observar el resultado, anoten ambos resultados. Luego repitan esta experiencia un número grande de veces 20 o 30 veces. Con la información obtenida discuta si los datos comprueban o desmienten la*

creencia de que el primer número que sale tiene menos oportunidad de salir favorecido en el segundo lanzamiento.

b) Una estudiante pregunta al docente ¿es cierto que las mandarinas regularmente tienen 10 gajos? El profesor que, realmente, nunca había prestado atención a eso; aprovechó la inquietud de la estudiante para introducir los conceptos de medidas estadísticas. En la siguiente lección trajo una mandarina para cada estudiante y les propuso que se la comieran, pero antes que contaran el número de gajos y el número de semillas de la fruta. Con ello, el docente, espera generar un grupo de datos que le permita introducir los conceptos sobre medidas estadísticas y al mismo tiempo evacuar la consulta de la estudiante, al menos para el grupo de mandarinas que consiguió.



Analice cada una de estas situaciones y responda:

- 1) ¿Qué técnica de recolección de información han utilizado los educadores en las experiencias anteriores?
- 2) En la actividad a) ¿cree usted que la experiencia planteada ayude a romper con la creencia tradicional de la falacia del jugador?
- 3) ¿Considera usted que la actividad b) provoca que se desaproveche mucho tiempo y genera un costo innecesario, debido a que el docente debió buscar o inventar otros datos que le ayudaran a introducir los conceptos de medidas estadísticas?

## Análisis de la Actividad 5

- 1) En las dos actividades propuestas por los docentes, se utilizan pequeños *experimentos* que ayudan a los estudiantes a buscar respuesta a un problema particular. Aunque se debe aclarar que en ambos casos la *experimentación* se combina con la *observación*, para generar los datos.
- 2) No se puede tener la certeza absoluta de que esta actividad rompa con dicha creencia, depende de varios factores, entre ellos la cantidad de repeticiones del experimento (entre más mejor). No obstante, al simular esta experiencia se espera que el número que salga favorecido en el primer lanzamiento, se repita una cantidad de veces similar a la de los otros cinco números para el segundo lanzamiento, y con ello, dejar claro que no existe una razón científica para creer que su probabilidad disminuye para el segundo lanzamiento. Con experimentos de este tipo, es posible romper algunas creencias que pueden afectar el aprendizaje de los nuevos conceptos.

- 3) En cuanto a la estrategia planteada por el docente para la segunda actividad, es de mucho valor práctico para la búsqueda de información dentro del contexto estudiantil, pues ante una pregunta simple, se desarrolla una experiencia que permite generar datos que ayudan a dar una respuesta a la interrogante; por lo que le da mayor realce a la disciplina pues el estudiante puede valorarla como una estrategia que le permite a comprender mejor el entorno.

En términos generales, la experimentación debe ser una herramienta fundamental para generar datos dentro del sistema educativo. Los niños y los jóvenes tienen un gran potencial para la experimentación, la técnica aprovecha ese potencial y lo encausa hacia el aprendizaje. Esta es una de sus principales ventajas; pero también pone al estudiante de cara a situaciones reales, y ante la búsqueda de estrategias que le permitan argumentar si una hipótesis o creencia está en lo correcto o no lo está. Su principal desventaja es que el docente debe realizar una muy buena planificación de los experimentos, e incluso ponerlos a prueba para tener una mayor certeza de lograr los propósitos educativos.

Por lo general, los experimentos se repiten varias veces para obtener suficientes datos que permitan llegar a conclusiones adecuadas y bien argumentadas.

## Actividad 6

*Debido a que se han presentado muchos problemas digestivos entre los estudiantes de un colegio, se desea realizar un estudio para identificar el tipo de alimentos que consumen los jóvenes. Para ello se elaboró se asignó la siguiente actividad.*

*En el artículo denominado, La dieta equilibrada: guía para enfermeras de Atención Primaria, de la Sociedad Española de Dietética y Ciencias de la Nutrición, publicada en la página Web [www.nutricion.org/publicaciones/pdf/Guía%20AP-DietéticaWeb.pdf](http://www.nutricion.org/publicaciones/pdf/Guía%20AP-DietéticaWeb.pdf), se incluye la siguiente información:*

### **INTRODUCCIÓN: LA DIETA EQUILIBRADA.**

*Alimentarse correctamente es una preocupación cada día más extendida en todos los países europeos. No solo es un tema que preocupe a los estudiosos, a los científicos y a los medios de comunicación, sino que muchas otras personas demuestran a diario su interés en seguir una alimentación equilibrada como un medio, el más sencillo, de alcanzar y disfrutar del bienestar personal. Numerosas enfermedades tienen una estrecha relación con nuestra dieta. Algunas pueden deberse a la escasez de ciertos nutrientes. Otras (y esto es mucho más frecuente en países como el nuestro) se deben a lo contrario: a la abundancia de alimentos que se suele aunar con la escasa actividad física, con el sedentarismo, y conducirnos a la obesidad y a otras patologías asociadas como es el caso de ciertos tipos de diabetes, hipertensión, alteraciones de los lípidos sanguíneos, etc.*

#### **- La alimentación equilibrada**

*La dieta equilibrada es aquella manera de alimentarse que aporta alimentos variados en cantidades adaptadas a nuestros requerimientos y condiciones*

personales. Llevar una alimentación equilibrada no es ingerir mucha comida, ya que es tan importante la cantidad como la calidad de la misma.

Es importante pensar que no se puede disfrutar de una vida saludable comiendo unos pocos alimentos. La **variación** es lo idóneo desde el punto de vista del bienestar, permitiéndonos esto disfrutar, además, de la riqueza de los distintos sabores, olores y aromas. De hecho, comer y beber forma parte de la alegría de vivir y, desde luego, la gastronomía no está reñida con las recomendaciones dietéticas ni con una alimentación saludable.

...  
Cuando nos referimos a los alimentos a lo largo de esta guía nos referimos a



La **Rueda de los alimentos** nos ayuda a conocer la proporción y la frecuencia de los alimentos en la dieta. De su forma se deduce que los alimentos que se encuentran en las 'porciones' o sectores más grandes (carbohidratos, frutas, verduras), son los que se deben ingerir en mayor cantidad. Si el sector es más pequeño, menor es la ingesta de los alimentos que debe provenir de los alimentados allí representados.

Vemos a continuación una sugerencia semanal de frecuencia de consumo de los alimentos más destacados. La presentación culinaria y la forma de cocinado deben de huir de la monotonía consiguiendo platos sabrosos y atractivos.

Las chucherías, bollos, refrescos azucarados, etc. deben de ser objeto de un consumo excepcional, no una costumbre. En su cocina utilice preferiblemente aceite de oliva. La cantidad total de alimentos a consumir depende de cada persona y sus características. Estas recomendaciones están realizadas para personas sanas.

## Consumo recomendado de alimentos



Con la intención de determinar el tipo de nutrición que tienen los estudiantes, se desea realizar un análisis estadístico que permita valorar la nutrición de los estudiantes del colegio. En este sentido, los estudiantes de la sección deben recolectar información que les permita realizar dicho análisis. Para ello pueden tomar en cuenta el siguiente cuestionario, como un apoyo para la recolección de la información requerida.

### Cuestionario

Estimado estudiantes, de la siguiente lista de productos, escriba una x en la celda que corresponda al nivel de consumo que usted realiza. ("Muy frecuente" corresponde a un consumo diario, "frecuente" corresponde a un consumo de entre cinco y seis veces por semana, "Regular" representa un consumo de tres o cuatro veces por semana, "Poco frecuente" representa un consumo de una o dos veces por semana, y "Rara vez" menos de una vez a la semana.)

Productos	Muy frecuente	Frecuente	Regular	Poco frecuente	Rara vez
1. Cereales, arroz, pasta, pan					
2. Hamburguesas, perros calientes, tacos, tortas, empanadas					
3. Frutas					
4. Verduras y ensaladas					
5. Papas fritas o tostadas, meneítos, picaritas, doraditas					
6. Carne, pescado, pollo, huevo					
7. Golosinas: caramelos, chocolates, gomitas, helados					
8. Lácteos: leche, queso, yogurt,					
9. Refrescos naturales o agua					

Muchas gracias por su colaboración

## Análisis de la Actividad 6

La técnica estadística que se ha utilizado en este caso corresponde al empleo de un *cuestionario*, que consistió en una serie de preguntas de respuesta única, que se redactaron utilizando una escala: *Muy frecuente, frecuente, Regular, Poco frecuente y Rara vez*. Para aplicarlo a los estudiantes del grupo.

Aunque el procedimiento utilizado es muy simple, y posiblemente para realizar un verdadero análisis de la problemática que atraviesa la institución se requiera de un estudio mucho más científico, lo que es cierto es que los datos que se puedan recabar por este medio permitirán tener una primera impresión sobre la magnitud del problema y formar conciencia en los estudiantes de lo conveniente de tener una adecuada alimentación.

## Encuestas

La técnica del cuestionario utilizada en la actividad anterior está incluida dentro de una estrategia más amplia que se denomina **encuesta**. Este es uno de los procedimientos más utilizados en estudios relacionados con el ámbito educativo. Las encuestas se caracterizan por plantear interrogaciones verbales o escritas dirigidas a las personas con la finalidad de recolectar datos para un estudio particular. Cuando la encuesta es verbal se hace uso de la técnica de la **entrevista**, pero en el caso escrito se le denomina **cuestionario**, que consiste en un listado de preguntas preestablecidas.

Una encuesta puede ser estructurada, cuando se compone de una serie de preguntas sistemáticamente elaboradas que se les formula a todos los encuestados de la misma forma. También puede ser no estructurada, cuando lo que se establece es únicamente una guía y el entrevistador puede modificar o elaborar nuevas dependiendo de las respuestas que se obtengan.

Las encuestas se le aplican a conjuntos de personas que poseen características similares, por ejemplo, las encuestas electorales que se realizan para medir la preferencia política entre los electores de una región o país.

## El cuestionario

Este instrumento se usa como estrategia de interrogación en diferentes formas: individual, grupal, telefónica, Internet, para registro de hechos, entre otros. Por su importancia, en este apartado se analizan algunos detalles claves en la elaboración de un cuestionario.

Las funciones fundamentales de este instrumento son las siguientes:

- 1) Obtener por medio de preguntas adecuadas, las respuestas que suministren los datos estadísticos necesarios para cumplir con los objetivos de la investigación. El cuestionario debe contener información pertinente, válida y confiable. Una pregunta es confiable si al ser aplicada en condiciones esencialmente iguales a los mismos individuos, en diversas oportunidades, da lugar a respuestas aproximadamente iguales. Una pregunta es válida si las respuestas a que da lugar reflejan exactamente la realidad que se desea medir.

2) Ayudar al entrevistador en la tarea de motivar al informante para que otorgue la información requerida. La redacción, el orden de las preguntas, el uso de palabras sencillas y claras, son fundamentales para lograr esa motivación en el informante.

Por otra parte, de la estructura y contenido del cuestionario depende, en gran medida, la calidad de la información que se obtenga. Los cuestionarios pueden ser de diversos tipos y formas dependiendo tanto de los propósitos de la investigación como de la población a la cual van dirigidos y de la técnica de recolección que se emplee (entrevista personal, telefónica, por correo).

Para realizar una adecuada entrevista debe hacerse un esfuerzo importante en localizar a la persona que será entrevistada, explicarle los objetivos del estudio y motivarla para lograr su cooperación.

En la elaboración del cuestionario debe dársele suma importancia a aspectos como: la escogencia de palabras (vocabulario adecuado y pertinente), el tipo de preguntas y la forma de hacerlas. Dentro de los tipos de preguntas se pueden mencionar las abiertas, cerradas y semi-abiertas y considerar las ventajas y desventajas de cada una de ellas.

Algunas de las consideraciones importantes para la construcción del cuestionario se describen seguidamente.

1) Incluir un pequeño párrafo de introducción donde se explique al entrevistado cuáles son los propósitos de la encuesta, qué instituciones lo están realizando y se le asegure la confidencialidad de la información.

2) Las respuestas deberían estar redactadas de manera que permitan una generación de datos sencillos, de modo que se puedan sistematizar y resumir de una manera simple.

3) Evitar incluir directamente las preguntas de la investigación en el cuestionario sin antes asegurarse que serán comprendidas por el entrevistado y contestadas adecuadamente.

4) Utilizar palabras precisas que aseguren la validez de las respuestas. Palabras claras y sencillas que signifiquen lo mismo para todos, que expresen claramente el concepto que se desea preguntar. Evitar las palabras difíciles o de uso poco común, las palabras regionales, porque en distintos lugares tienen significados diferentes.

5) Las preguntas no deben ser ambiguas, ni sugerir las respuestas.

6) No todas las preguntas se pueden incluir en el cuestionario en aras de la brevedad del mismo.

7) Colocar las preguntas en un orden conveniente. Al inicio las más fáciles de contestar o de carácter general y al centro o la final las más difíciles o de naturaleza muy personal.

8) Las pruebas previas del cuestionario permiten determinar el orden más adecuado para las preguntas.

9) Evitar las preguntas bidimensionales (preguntar dos cosas en una sola interrogante).

10) Evitar las preguntas en las que aparezca una negación, pues frecuentemente son mal interpretadas.

Por otro lado, para la aplicación del cuestionario se deberían considerar ciertos aspectos como los que se mencionan a continuación.

1) Escogencia adecuada de las personas que van a aplicar el instrumento (conducta, cortesía, relaciones con los demás).

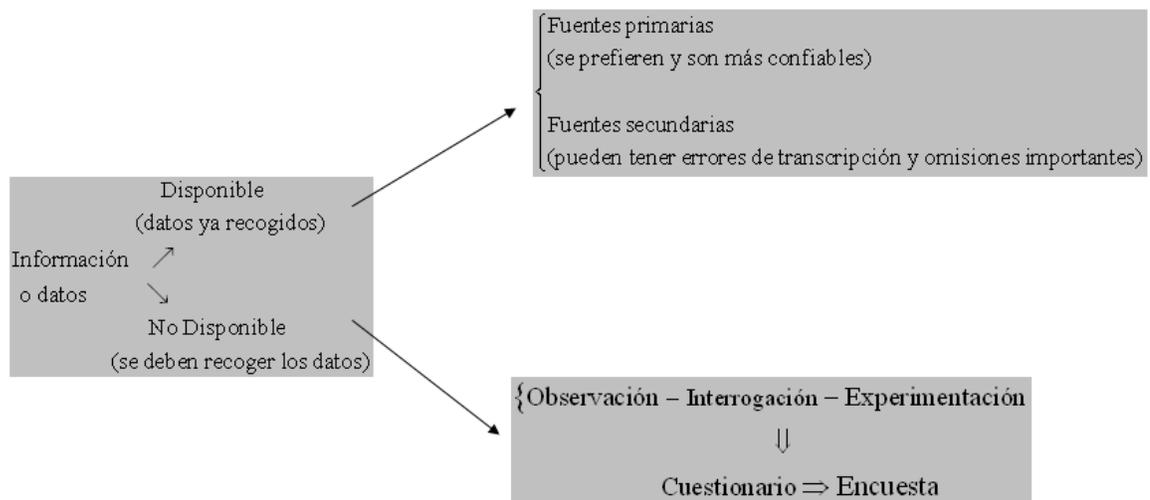
2) Instrucción y preparación de los encuestadores respecto al cuestionario y los propósitos del estudio.

3) Motivar al informante para que contribuya y suministre la información lo más confiable posible.

4) Revisión de la información recabada para verificar que no hay contradicciones.

Finalmente, una vez que se han completado los cuestionarios, la información que hay en ellos debe ser procesada para convertirla en cifras numéricas que puedan ser sometidas al análisis estadístico; y luego presentarla en forma resumida mediante cuadros, gráficas y medidas descriptivas, con el propósito de comunicar, interpretar y describir la situación real planteada.

A manera de resumen sobre el tópico recolección de información se presenta el siguiente esquema.



## IV. Representaciones

### Introducción

Al trabajar el área de la Estadística se utilizan diferentes herramientas que favorecen el análisis de los datos. Las representaciones tabulares y gráficas son algunas de las técnicas más utilizadas para ese fin. No obstante, en los procesos educativos se debe tener presente que la construcción de este tipo de representaciones no constituye el fin, sino un medio para comunicar y analizar información que permita extraer conclusiones relevantes para situaciones particulares.

Este apartado trata sobre las representaciones tabulares y gráficas, se abarca la construcción de las mismas y algunos lineamientos para una interpretación adecuada.

### Actividad 1

*Lea y analice la siguiente situación:*

*El director de una institución particular requiere saber a qué dedican el tiempo, fuera de clases, los estudiantes del III y IV Ciclo, para así, entre otras cosas, poder hacer comparaciones por ciclo y buscar una posible causa del bajo rendimiento.*

*Los aspectos a considerar para cada estudiantes serían: sexo, nivel en el que se encuentra, tipo de actividad realizada fuera del horario de clase (mirar la televisión, practicar algún deporte, realizar deberes escolares, leer, compartir con amigos fuera de su casa, navegar en la Web y uso de redes sociales, practicar juegos electrónicos o computacionales, realizar deberes domésticos, dormir, entre otros) y tiempo invertido en dicha actividad por día.*

*El director cuenta, solamente con una semana para realizar la investigación, no tiene experiencia en realizar muestreos aleatorios, por lo que decide construir un cuestionario y administrárselo al total de estudiantes en esos ciclos, dicha población es de 150 estudiantes.*

*Los resultados de dicha investigación serán analizados en conjunto con las calificaciones del trimestre, con el propósito de establecer algún tipo de relación entre el uso del tiempo fuera de clases y las calificaciones obtenidas. Se espera presentar a los padres de familia estos resultados cuando se entreguen los reportes de calificaciones, con el fin de que ellos puedan analizar esta información con sus hijos.*

*Los datos recolectados son de la forma que aparece a continuación:*

Estudiante	Nivel	Sección	Sexo	Número de horas invertidas para cada actividad									
				Deberes escolares	Televisión	Deporte	Leer	Amigos	Internet	Juego electrónico	Deberes domésticos	Dormir	Otros
A	sétimo	7.1	Mujer	1,0	3,0	0,0	0,0	2,0	0,0	0,0	4,0	6,0	2,0
B	sétimo	7.3	Hombre	2,0	3,5	2,0	0,0	1,0	1,0	0,0	1,0	6,5	1,0
C	octavo	8.2	Hombre	1,5	2,0	2,0	0,5	0,5	3,0	1,0	0,0	7,0	0,5
D	décimo	10.1	Hombre	1,0	2,5	0,0	1,0	1,5	3,0	2,0	1,0	6,0	0,0
E	octavo	8.2	Mujer	3,0	1,5	1,0	0,0	1,0	2,5	1,0	0,0	8,0	0,0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Si únicamente se cuenta con esta información, conteste las siguientes preguntas:

1. ¿Se pueden organizar de alguna manera los datos recolectados por variable, para tenerlos de manera resumida?
2. ¿De qué manera puede el director presentar la información a los padres de familia para que estos puedan entenderla, le presten atención y sean conscientes de la situación?

## Análisis de la Actividad 1

- 1) Al observar los datos que se recolectaron, puede sugerirse que para cada variable, se proceda a resumir la información determinando la frecuencia o número de ocurrencias para cada valor obtenido, por medio de este procedimiento se puede determinar el patrón de variación de los valores observados y tener así una idea básica sobre las actividades más importantes a las que los estudiantes dedican el tiempo fuera de clases.
- 2) Con la información anterior es posible buscar algunas técnicas estadísticas para la representación de la información, de este modo sería posible resumir y presentar los datos de una manera sencilla para que los padres de familia puedan asimilar el mensaje que se les desea comunicar. Por ejemplo podría elaborarse la siguiente tabla

*N° de horas que invierten los escolares por día a la realización de deberes escolares*

N° de horas	N° de estudiantes
1	67
2	23
3	14
4	2
⋮	⋮
Total	150

El ejemplo anterior, muestra la importancia de las técnicas estadísticas para resumir y presentar datos. Dentro de un estudio estadístico, una vez que los datos han sido recolectados; es necesario organizarlos de una manera sistemática que posibilite su resumen, y prepare el camino para una presentación textual, tabular o gráfica, según corresponda.

Las formas de presentación de información se eligen según el tipo de datos que se tenga, el mensaje que se desea comunicar, el tipo de lector al que va dirigido el mensaje y el efecto que se quiere provocar en ese lector.

## Actividad 2

*Un profesor necesita hacer un informe final acerca del rendimiento académico de los estudiantes de octavo año. Para ello cuenta con las siguientes variables vinculadas con los jóvenes: sexo, sección y las notas en Matemáticas del II Trimestre de las cinco secciones de octavo, en total son 154 estudiantes. Un ejemplo de los datos que tiene se adjunta seguidamente.*

Nombre	Sección	Sexo	Nota
Alpízar Salazar Mario	8-1	H	82
López Flores Xiomara	8-3	M	71
Mejía Araya Marlon	8-2	H	56
Solís Rodríguez Aaron	8-1	H	63
Soto Paris María Eugenia	8-5	M	87
⋮	⋮	⋮	⋮

Con base en la información suministrada conteste las siguientes preguntas:

1. ¿Qué variables se podrían comunicar de manera textual?
2. ¿Qué variables se podrían comunicar de manera tabular?
3. ¿Qué variables se podrían comunicar de manera gráfica?

## Análisis de la Actividad 2

1. La clasificación según el sexo sería un aspecto a presentar de manera textual, por ejemplo, *“En total se analizó la información de 48 mujeres y 106 hombres”*, aunque también se puede comunicar mediante porcentajes, indicando el porcentaje correspondiente a uno de los sexos, por ejemplo, *“De los estudiantes analizados en el estudio 31,2% son mujeres”*, no hace falta indicar que 68,8% eran hombres, pues esto lo puede deducir el lector fácilmente, hay que procurar no ser redundante.
2. En cuanto a las notas, resulta complejo presentarlas en forma textual pues debido a la gran variabilidad de datos, se requiere sistematizarlos en tablas estadísticas ya sea por agrupaciones de notas: por ejemplo, *menores de 40, entre 40 y menos de 65, de 65 a menos de 80, de 80 a menos de 90, 90 o más*; o cualquier otra agrupación.
3. Las notas también pueden presentarse mediante diagramas que permiten representar todas las notas en una misma representación gráfica, tal como será analizado, más adelante, en el documento.

## Presentación textual

Como se ilustró en la actividad 2, en algunas ocasiones la información que proporcionan los datos puede ser presentada en forma textual. En este tipo de presentación se introducen cifras dentro del texto. Al irse escribiendo el informe, se incluyen ciertas cifras que se consideran importantes para el estudio y así llamar la atención sobre ellas.

## Presentación tabular

Cuando se quiere introducir un mayor nivel de detalle de la información proporcionada por los datos, se puede proceder a la elaboración de un cuadro. Esta técnica corresponde a una sistematización de las observaciones en filas y columnas, de acuerdo a la clasificación de interés. Este resumen, por medio de cuadros, debe facilitar la interpretación de la información que comunican estos datos, y ser suficientemente explícita, de modo que el lector no deba recurrir al texto para comprender el cuadro.

Desde un punto de vista didáctico, la enseñanza de la Estadística, no debe centrarse en los procesos de construcción de cuadros, pues ellos sólo son una herramienta que apoya el proceso de resumen de los datos para facilitar su interpretación. No obstante, debe analizarse si la estructura del cuadro construido permite a un lector comprender la información que se quiere transmitir.

## Representación tabular

### Actividad 3

*A continuación se presenta una base de datos ficticia, con fines didácticos.*

*Las autoridades del Hospital Calderón Guardia necesitan caracterizar la población que asiste a la sección de emergencias los fines de semana, para ello deciden hacer una recolección de datos en un fin de semana aleatoriamente seleccionado de cada mes.*

*Para recolectar los datos, lo primero que hacen es pasar a cada paciente a la sala de enfermería, ahí una enfermera toma la presión arterial, mide su estatura y peso. Luego pasan a ventanilla y deben llenar una hoja de registro donde se les pide: sexo, lugar de residencia, edad y el número de veces que visitó el servicio de emergencias en el 2010.*

*La siguiente información pertenece al tercer fin de semana del mes de abril del año 2011.*

Asegurado	Lugar de residencia	Sexo	Visitas emergencias en el año 2010	Presión arterial	Edad (años)	Peso (kg)
1	Coronado	M	4	N	38	61,0
2	Moravia	M	6	N	46	55,0
3	Goicoechea	M	5	A	29	79,1
4	Moravia	H	2	B	38	70,7
5	Montes de Oca	M	6	A	46	70,8
6	Montes de Oca	M	4	N	44	55,9
7	Coronado	M	4	N	63	72,2
8	San José	H	3	A	56	75,1
9	Moravia	M	1	N	43	60,8
10	Goicoechea	M	5	N	54	70,6
11	San José	M	2	N	72	71,6
12	Goicoechea	H	1	B	71	85,5
13	Moravia	M	1	N	51	73,6
14	Goicoechea	H	3	B	23	75,2
15	Montes de Oca	M	1	N	45	78,5
16	Montes de Oca	H	2	N	60	75,0
17	Montes de Oca	H	2	A	37	79,1
18	San José	M	6	A	40	68,2
19	Goicoechea	M	1	N	49	76,2
20	Goicoechea	M	1	N	51	50,4
21	Coronado	H	2	N	47	88,8
22	Moravia	M	3	B	53	78,8
23	Moravia	M	4	N	40	48,7
24	San José	M	5	A	48	54,7
25	Tibás	H	2	N	54	82,1
26	San José	M	4	N	28	56,6
27	Coronado	M	3	N	34	65,7
28	Tibás	H	1	N	44	75,4
29	San José	H	3	A	34	78,2
30	San José	M	2	N	35	76,9
31	Tibás	H	2	B	58	80,1
32	Moravia	H	1	A	31	82,5
33	Moravia	H	3	N	16	72,6
34	Moravia	H	4	N	43	67,7
35	San José	M	5	N	27	65,8
36	Goicoechea	H	1	N	59	94,9
37	Tibás	H	1	N	84	92,6
38	Tibás	M	2	N	66	48,1
39	San José	M	5	A	39	50,5
40	Montes de Oca	H	2	A	21	70,2

En la casilla de sexo: M significa Mujer y H hombre.  
 En la casilla de presión arterial: B significa baja, N normal y A alta  
 La edad se midió en años cumplidos.

Con base en los datos recopilados:

1) Construya un cuadro donde se agrupe y resuma el lugar de residencia de los pacientes. De acuerdo con los datos del cuadro construido, conteste las siguientes preguntas:

- ¿En cuál cantón vive la mayor parte de los pacientes?
- ¿El lugar de residencia de los pacientes es variable o se concentra alrededor de un cantón?

2) Construya un cuadro donde se resuma el número de veces que ha visitado emergencias en el año 2010. Basados en el cuadro construido, conteste las siguientes preguntas:

- ¿Qué porcentaje de personas ha asistido al servicio de emergencias en exactamente tres ocasiones?
- ¿Qué porcentaje de personas ha asistido al servicio de emergencias en al menos tres ocasiones?
- ¿Cuántos pacientes entrevistados en el tercer fin de semana de abril del 2011, asistieron al Servicio de Emergencias entre dos veces y cuatro veces (inclusive) durante el 2010?

3) Por último, acerca de la frecuencia con que los entrevistados visitan el servicio de emergencias se puede observar que la mayoría lo ha hecho menos de tres veces en el año.

Construya un cuadro donde se resuma la edad de los pacientes, de acuerdo con los datos de este cuadro, conteste las siguientes preguntas:

- ¿Cuál es el intervalo de edad donde se presenta la mayor frecuencia de pacientes?
- Dé una interpretación general de la información del cuadro.

### Análisis de la Actividad 3

1) Si la variable a representar es de naturaleza cualitativa, como es el caso del lugar de residencia, la construcción del cuadro es muy simple en el sentido de que se caracterizan por columna o fila según se requiera. En este caso el cuadro consta de dos columnas, en la primera se colocan los cantones y en la segunda la cantidad de pacientes que habitan cada cantón; es importante colocar encabezados de columnas contextualizados y un título, como se muestra a continuación.

**Cuadro 1. Lugar de procedencia de los pacientes que asistieron a Emergencias del Hospital Calderón Guardia, en el tercer fin de semana del mes de abril, 2011**

Lugar de residencia	Cantidad de pacientes
San José	9
Moravia	9
Goicoechea	7
Montes de Oca	6
Tibás	5
Coronado	4
<b>Total</b>	<b>40</b>

Hay dos cantones que presentan la mayor frecuencia de asistencia a Emergencias de ese hospital, a saber San José y Moravia (9 pacientes). Sin embargo, los demás cantones también presentan una frecuencia alta, por ello se refleja también una alta variabilidad en el lugar de residencia de pacientes respecto a estos seis cantones

- 2) Si la variable es cuantitativa discreta y el número datos diferentes es relativamente pequeño, para la construcción de un cuadro se puede indicar la frecuencia de cada uno de ellos, de una forma similar a la que se realiza con las categorías para una variable cualitativa. En este caso, la cantidad de visitas al servicio de emergencias varían entre una y seis, por ello estos números se incluyen en la primera columna mientras que la frecuencia con la que fueron los pacientes va en la segunda, como se muestra en el cuadro adjunto.

**Cuadro 2. Cantidad de veces que los pacientes que se presentaron a la consulta de Emergencias del Hospital Calderón Guardia en el tercer fin de semana del mes de abril del 2011, visitaron esa consulta en el 2010**

Número de veces	Cantidad de pacientes
1	10
2	10
3	6
4	6
5	5
6	3
<b>Total</b>	<b>40</b>

Según los datos del cuadro anterior seis de las cuarenta personas asistieron exactamente tres ocasiones al servicio de emergencias en el 2010 esto equivale a 15,0%. No obstante, una forma más práctica de evidenciar el peso relativo de cada valor, consiste en incluir una columna a la derecha de estos cuadros donde se incluye su valor porcentual. En este sentido, los cuadros 1 y 2 quedarían de la siguiente forma.

**Cuadro 1. Lugar de procedencia de los pacientes que asistieron a Emergencias del Hospital Calderón Guardia, en un fin de semana del mes de abril, 2011**

Lugar de residencia	Cantidad de pacientes	Porcentaje de pacientes
San José	9	22,5
Moravia	9	22,5
Goicoechea	7	17,5
Montes de	6	
Oca		15,0
Tibás	5	12,5
Coronado	4	10,0
<b>Total</b>	<b>40</b>	<b>100,0</b>

**Cuadro 2. Cantidad de veces que los pacientes que se presentaron a la consulta de Emergencias del Hospital Calderón Guardia en el tercer fin de semana del mes de abril del 2011, visitaron esa consulta en el 2010**

Número de veces	Cantidad de pacientes	Porcentaje de pacientes
1	10	25,0
2	10	25,0
3	6	15,0
4	6	15,0
5	5	12,5
6	3	7,5
<b>Total</b>	<b>40</b>	<b>100,0</b>

La importancia de incluir los porcentajes es que le permite al lector determinar el peso relativo de cada dato. En este sentido, si se desea conocer el porcentaje de pacientes que asistió al servicio de Emergencias en al menos tres veces (significa tres o más) se suman los porcentajes correspondientes a los valores tres, cuatro, cinco y seis, es decir  $15,0 + 15,0 + 12,5 + 7,7 \approx 50\%$ . Esto indica que la mitad de los pacientes asistió a este servicio en al menos tres ocasiones durante el 2010.

Por otro lado, para conocer la cantidad de pacientes que asistió al servicio de emergencias entre dos y cuatro veces (inclusive) durante el 2010, se deben sumar los que asistieron dos, tres y cuatros veces, ellos son  $10 + 6 + 6 = 22$ , esto equivale a un  $25,0 + 15,0 + 15,0 = 55\%$ . Es decir 55% de los pacientes, asistió en el 2010 al servicio de Emergencias de ese hospital entre dos y cuatro veces (inclusive).

**Nota:** Al ser este valor mayor al 50%, se puede decir que la mayoría de los pacientes atendidos en este servicio en el tercer fin de semana del mes de abril, había asistido entre dos y cuatro veces durante el año 2010.

3) La variable edad, al estar vinculada con el tiempo es una variable continua; pero si únicamente se considera en años cumplidos entonces se convierte en variable discreta (es contable). A pesar de esto, por la gran variabilidad de las edades observadas, resulta monótono construir un cuadro similar a los anteriores, por esta razón se recurre a estrategias de agrupación para distribuir las edades y resumir la información de una mejor manera. Un procedimiento muy utilizado consiste en determinar el dato menor y el mayor para luego establecer categorías o grupos de datos, preferiblemente del mismo tamaño que incluyan todas las observaciones. Por ejemplo, en este caso la edad más baja es 16 y la más alta 84 años, entonces se pueden establecer diferentes tipos de agrupaciones, una de ellas podría ser iniciar en 15 años y agrupar las edades cada 10 años, por ejemplo los grupos serían “de 15 a menos de 25”, “de 25 a menos de 35”, y así sucesivamente, hasta llegar al grupo “de 75 a menos de 85”, como se presenta en el siguiente cuadro.

**Cuadro 3. Distribución de las edades de los pacientes que asistieron al servicio de Emergencia del Hospital Calderón Guardia durante el tercer fin de semana del mes de abril del 2011**

Edad de los Pacientes (años)	Cantidad de pacientes	Porcentaje de pacientes
De 15 a menos de 25	3	7,5
De 25 a menos de 35	6	15,0
De 35 a menos de 45	11	27,5
De 45 a menos de 55	11	27,5
De 55 a menos de 65	5	12,5
De 65 a menos de 75	3	7,5
De 75 a menos de 85	1	2,5
<b>Total</b>	<b>40</b>	<b>100</b>

Nuevamente, los encabezados de columnas contextualizados y el título del cuadro son elementos indispensables para la correcta interpretación del mismo.

Existen dos intervalos que son los de mayor frecuencia, ellos recogen edades entre los 35 y 55 años, cada una de estas clases incluye 11 pacientes, entre ambos acumulan 55% de los pacientes que llegaron ese fin de semana.

Al momento de interpretar la información de un cuadro (gráfica o cualquier otra figura), no se debe repetir los datos que el mismo proporciona, sino hacer una lectura más allá de los datos, por ejemplo, en este caso algunos de los aspectos que se pueden resaltar respecto a las edades de los pacientes que asistieron en el tercer fin de semana de abril del 2011 al servicio de Emergencias del Hospital Calderón Guardia son:

- Un alto porcentaje (70%) de los pacientes atendidos en ese fin de semana corresponde a personas adultas que van entre los 25 y los 55 años. Además, más de la mitad (55%) tenía edades entre los 35 y los 55 años.
- Únicamente, se atendieron tres adultos mayores en ese fin de semana (10%).
- Se presentó gran variabilidad entre las edades de los pacientes atendidos, lo cual es lógico dentro de este tipo de servicios de emergencia.
- No se atendieron niños y únicamente, se atendieron tres adolescentes.

Estos son algunos de los elementos que se pueden citar. Para efectos de realizar otros análisis habría que tener presente los objetivos que originaron el estudio que llevó a la recolección de esa información.

**Nota:** Observe que la forma de resumen utilizada para el último cuadro, no sigue un patrón único, sino que se pudieron utilizar otras agrupaciones que también podrían ser válidas, como las que siguen:

Edad de los pacientes	Número de pacientes	Edad de los pacientes	Número de pacientes
De 15 a menos de 20	1	Menos de 20 años	1
De 20 a menos de 25	2	De 20 a menos de 30	5
De 25 a menos de 30	3	De 30 a menos de 40	8
De 30 a menos de 35	3	De 40 a menos de 50	12
De 35 a menos de 40	5	De 50 a menos de 60	8
De 40 a menos de 45	6	De 60 a menos de 70	3
De 45 a menos de 50	6	De 70 a menos de 80	2
De 50 a menos de 55	5	De 80 a menos de 85	1
De 55 a menos de 60	3	<b>Total</b>	<b>40</b>
De 60 a menos de 65	2		
De 65 a menos de 70	1		
De 70 a menos de 75	2		
De 75 a menos de 80	0		
De 80 a menos de 85	1		
<b>Total</b>	<b>40</b>		

Con estas otras representaciones la interpretación varía ligeramente; pero, los principios básicos del análisis se mantienen.

## Observaciones generales

Basados en las representaciones construidas y las respuestas a las preguntas planteadas:

¿Existe alguna diferencia en la construcción de los cuadros anteriores? Justifique su respuesta.

Los cuadros anteriores poseen diversas diferencias y similitudes, entre las semejanzas están:

- Todos tienen un título que permite al lector comprender la información que suministra.
- Todos incluyen la columna matriz que suministra información de los datos de las filas, además contienen la fila en la que se incluyen los encabezados de las columnas, que suministran el tipo de información de las columnas transmiten.
- En todos se incluye frecuencia o número de pacientes relacionada con cada categoría, número o agrupación. Pero además, también se ha visto la necesidad de incluir porcentajes para valorar el peso relativo de cada uno de ellos.

Entre las diferencias se tiene la naturaleza de la variable y la variabilidad de los datos, lo que puede conducir a realizar procedimientos de agrupación de datos en aquellos casos en lo que la variabilidad de las respuestas es tan grande que provoca que se generen muchas observaciones muy diferentes entre sí.

\*\*\*

En general, para que la información desplegada en un cuadro llegue al lector de la mejor manera, ellos deben contener: título, columna matriz, encabezados contextualizados, cuerpo o contenido, y en los casos en los que la información ha sido tomada de una fuente primaria o secundaria se debe indicar esa fuente.

**Título:** Respecto al título, debe explicar: ¿Cuál es la naturaleza de los datos o a qué hacen referencia? ¿Dónde se recogieron los datos o a qué lugar corresponde? ¿Cuándo ocurrieron los hechos a que se refieren los datos?, además; ¿cuál es la clasificación o clasificaciones principales empleadas para organizar los datos?

**Encabezados:** Los encabezados deben ser claros y acordes con el contexto analizado. El contenido del cuadro debe ser explícito y coherente en con el título y los encabezados.

**Fuente:** La fuente debe incluir, siempre que sea posible, todos los requisitos de una cita bibliográfica: autor, año, nombre de la publicación, editorial, cuadro y página de dónde proceden los datos.

A continuación se da un ejemplo de un cuadro estadístico, con todos los elementos sugeridos:

**Cuadro 4.28**

**Recibió algún curso en el que se impartieran los temas de Probabilidad y Estadística.  
DREH y DREPZ. Año 2009**

Recibió	Cantidad de docentes				
	Heredia	Pérez Zeledón	Total		
Sí	73 (55,7)	39 (70,9)	112 (60,2)		
No	58 (44,3)	16 (29,1)	74 (39,8)		
<b>Total</b>	<b>131 (100,0)</b>	<b>55 (100,0)</b>	<b>186 (100,0)</b>		

21 educadores no respondieron.

**Nota aclaratoria**

Fuente: Barrantes et al. (2010). Memoria de seminario "Estado actual de la enseñanza y aprendizaje de probabilidad y Estadística, en I y II Ciclo, en la Educación costarricense en las direcciones regionales educativas de Heredia y Pérez Zeledón". Universidad Nacional, Heredia, Costa Rica, Cuadro 4.28, página 242.

Otros aspectos que deben tomarse en consideración para la construcción de cuadros son: tamaño, formato de las cantidades que van dentro (igual número de decimales), bordes (solamente en los encabezados y al final para cerrar el cuadro) y el ordenamiento de las categorías en los encabezados (alfabético, por magnitud, cronológico, geográfica).

Como se citó anteriormente, el objeto de estudio de la Estadística no es la construcción de representaciones sino la interpretación y análisis de los datos, por lo que es de suma importancia dar énfasis a la información que se pueda extraer de una representación tabular.

## Distribuciones de frecuencia

El nombre que comúnmente reciben los cuadros donde son clasificadas variables estadísticas, ya sean cuantitativas o cualitativas es *distribución de frecuencias*. Estas representaciones ayudan a analizar la forma o patrón de la distribución de los datos (simétrica, sesgada), la posición de la distribución, es decir, alrededor de qué valor se tienden a concentrar los datos (valores centrales) y la dispersión o variabilidad de ellos alrededor de los valores centrales.

Para agrupar los datos se dividen en clases o categorías, dentro de lo posible todas del mismo tamaño. Se sugiere utilizar más de cinco y menos de 15 agrupaciones, y esta cantidad depende del número total de observaciones. En variables cuantitativas continuas deben evitarse las clases de diferente amplitud y también las clases abiertas. Además, se prefiere, hasta donde sea posible, que el intervalo de clase sea cinco, diez, o un múltiplo de ellos.

Dichos cuadros deben contener en la primera columna (de izquierda a derecha) las clases: categorías utilizadas cuando se trabaja con variables cualitativas; los intervalos o valores discretos cuando se trata de variables cuantitativas.

Si se trabaja con variables cuantitativas continuas se deben delimitar los límites de clase que son los valores que definen una clase separándola de la anterior y de la posterior. Deben ser tales que permitan definir clases exhaustivas (clasificar a todas las

observaciones en alguna de ellas) y mutuamente excluyentes (no permitir que una observación quede incluida en más de una clase). Para definir el límite inferior de la primera clase se considera el dato menor del conjunto de observaciones y se escoge un valor más pequeño a éste.

En la segunda columna se puede colocar la frecuencia absoluta de cada clase o intervalo, que es el número de veces que aparece un dato. Esta columna debe sumar el número total de observaciones.

En la tercera columna la frecuencia porcentual que representa el porcentaje de datos en cada clase (también puede ser la proporción, denominada frecuencia relativa, que son los cocientes entre las frecuencias absolutas de cada clase y el número total de datos).

## Actividad 4

*Analice la información proporcionada en el siguiente cuadro y con base en ella responda las preguntas que se formulan luego.*

**Cuadro 4. Estimación de la población de 70 años y más para Costa Rica al 30 de junio del 2012, según el sexo**

Edad en años cumplidos	Hombre	Mujer
70-74 años	39 163	42 686
75-79 años	27 026	32 285
80-84 años	17 675	22 656
85-89 años	8 937	12 394
90-94 años	3 230	4 853
95 años y más	1 207	1 930
<b>Total</b>	<b>97 238</b>	<b>116 804</b>

Fuente: Base de datos electrónica del Centro Centroamericano de Población-UCR  
Dirección Web: [ccp.ucr.ac.cr](http://ccp.ucr.ac.cr)

- 1) *Determine los porcentajes de población, por sexo, para cada uno de los grupos de edad.*
- 2) *¿Qué información relevante puede resaltar de esos datos en relación con los porcentajes de personas para cada sexo? ¿Qué se puede concluir?*

## Análisis de la Actividad 4

- 1) Para responder lo que se solicita, es necesario agregar dos nuevas columnas al cuadro, que pueden ser de la siguiente manera.

**Cuadro 4. Estimación de la población de 70 años y más para Costa Rica al 30 de junio del 2012, según el sexo**

Edad en años cumplidos	Absolutos		Porcentajes	
	Hombre	Mujer	Hombre	Mujer
70-74 años	39 163	42 686	40,3	36,5
75-79 años	27 026	32 285	27,8	27,6
80-84 años	17 675	22 656	18,2	19,4
85-89 años	8 937	12 394	9,2	10,6
90-94 años	3 230	4 853	3,3	4,2
95 años y más	1 207	1 930	1,2	1,7
<b>Total</b>	<b>97 238</b>	<b>116 804</b>	<b>100</b>	<b>100</b>

Fuente: Base de datos electrónica del Centro Centroamericano de Población-UCR  
Dirección Web: [ccp.ucr.ac.cr](http://ccp.ucr.ac.cr)

- 2) Acá hay que hacer notar que los porcentajes calculados se convierten en una herramienta fundamental para el análisis de la segunda pregunta. En general, se puede notar que se ha proyectado que lleguen a estas edades más mujeres que hombres ( $116\ 804 - 97\ 238 = 19\ 570$ ); es decir, en esas edades, se ha proyectado que lleguen con vida cerca de 20 000 mujeres más que hombres. Si todavía se ve más allá, puede notarse que del total de personas que se espera lleguen con vida a los 70 o más años en el 2 012 (214 042) 54,6% son mujeres. Pero además, aunque los porcentajes de ambos grupos descienden conforme avanza la edad, se observa que son más altos entre los hombres en edades menores y más bajos en las edades mayores. En general, una hipótesis de estos resultados puede ser que, para grupo de edad, se presentan más defunciones de hombres que de mujeres.

En este ejemplo se demuestra que el lector más que leer los datos debe ver el mensaje que hay detrás de ellos. En este sentido, se debería promover que el ciudadano sea capaz de realizar estas interpretaciones y vincularlas con la información que normalmente se difunde en los medios de comunicación. Por ejemplo, es bien conocido que la esperanza de vida al nacer (o promedio de vida esperada) es más alta entre las mujeres respecto a los hombres, entonces los resultados citados en el cuadro anterior son un reflejo de este resultado; por lo que la hipótesis dada al final del análisis resulta acertada en la vida real.

\*\*\*

La representación anterior es muy utilizada para relacionar dos tipos de variables, este caso se tenía la distribución de población por edad y por sexo. Estas representaciones se denominan *cuadros de datos cruzados o de doble entrada*. Otro ejemplo, de este tipo de cuadros es el siguiente:

## Actividad 5

Se desea analizar la relación entre el deporte preferido y sexo para los estudiantes del Ciclo Diversificado del Colegio Dominicano.

**Cuadro 5. Distribución de los estudiantes del Ciclo Diversificado del Colegio Dominicano, según el deporte preferido y el sexo**

Deporte favorito	Mujeres	Hombres	Número de estudiantes
Fútbol	16	31	47
Basquetbol	12	13	25
Beisbol	5	8	13
Natación	4	6	10
Ciclismo	3	4	7
Otro	2	4	6
<b>Total</b>	<b>42</b>	<b>66</b>	<b>108</b>

Fuente: Encuesta realizada por el profesor de Educación Física del Colegio Dominicano.

Utilice la información para llevar a cabo un análisis comparativo entre la preferencia por el deporte y el sexo del estudiante.

## Análisis de la Actividad 5

Como puede notarse, los datos absolutos no permiten llevar a cabo una comparación objetiva por sexo debido a que los totales relativos para cada categoría son muy diferentes, por esta razón se requiere determinar *datos porcentuales*. Debido a que la comparación que se desea hacer es por sexo, entonces los porcentajes deben calcularse para cada categoría para el total del sexo, tal como se muestra en el cuadro que sigue:

**Cuadro 5. Distribución de los estudiantes del Ciclo Diversificado del Colegio Dominicano, según el deporte preferido y el sexo.**

Deporte favorito	Mujeres		Hombres	
	Absoluto	Porcentaje	Absoluto	Porcentaje
Fútbol	16	38,1	31	47,0
Basquetbol	12	28,6	13	19,7
Beisbol	5	11,9	8	12,1
Natación	4	9,5	6	9,1
Ciclismo	3	7,1	4	6,1
Otro	2	4,8	4	6,1
<b>Total</b>	<b>42</b>	<b>100</b>	<b>66</b>	<b>100</b>

Fuente: Encuesta realizada por el profesor de Educación Física del Colegio Dominicano.

Este nuevo cuadro refleja que las preferencias deportivas presentan algunas diferencias entre los hombres y las mujeres, en particular los hombres tienen mayor preferencia por el fútbol que las mujeres, mientras que ellas tienen una mayor preferencia por el básquetbol. En el resto de disciplinas no hay grandes diferencias. Tampoco se observan diferencias importantes en el patrón de preferencias.

\*\*\*

No siempre el uso de porcentajes es adecuado para favorecer una interpretación, en el siguiente ejemplo los datos absolutos bastan para comprender el comportamiento de la situación planteada y los porcentajes pueden traer alguna confusión.

## Actividad 6

Observe el siguiente cuadro correspondiente a información de un colegio, la cual ha sido adaptada con fines didácticos.

**Cuadro 6. Distribución de las calificaciones por trimestre obtenidas por el nivel de octavo año del colegio hipotético, 2010**

Calificaciones	I trimestre		II Trimestre		III Trimestre	
	Total	Porcentaje	Total	Porcentaje	Total	Porcentaje
De 15 a menos de 30	5	6,3	1	1,6	0	0,0
De 30 a menos de 45	10	12,5	4	6,6	0	0,0
De 45 a menos de 70	30	37,5	22	36,1	18	36,0
De 70 a menos de 85	25	31,3	24	39,3	22	44,0
De 85 a 100	10	12,5	10	16,4	10	20,0
<b>Total</b>	<b>80</b>	<b>100,0</b>	<b>61</b>	<b>100,0</b>	<b>50</b>	<b>100,0</b>

Fuente: Dirección Administrativa del colegio hipotético, 2010.

Con base en la información del cuadro conteste las siguientes preguntas:

- 1) ¿Qué puede decir acerca de la deserción de los estudiantes de octavo año del Colegio Hipotético (realice el análisis comparando los trimestres)?
- 2) ¿Qué puede decir acerca del rendimiento académico de los estudiantes de octavo año del Colegio Hipotético (puede hacer el análisis por trimestres)?

## Análisis de la Actividad 6

Si únicamente se observan los porcentajes se podría creer que el rendimiento académico ha venido aumentando paulatinamente a medida que avanza el año, no obstante al analizar los datos absolutos la situación cambia drásticamente.

- 1) Mediante el análisis de los valores totales se observa una fuerte deserción, de modo que de los 80 que concluyeron el primer trimestre, únicamente 50 concluyen las evaluaciones del tercer trimestre. No se puede cuantificar la deserción total pues se desconoce cuál era la matrícula original de este nivel educativo; pero, si se utiliza como base el total de estudiantes que concluyó el primer trimestre, se tiene 19 estudiantes desertaron o se trasladaron durante el segundo trimestre y 11 lo hicieron durante el tercer trimestre. Con esta información se tiene que 19 de 80 representa 23,8% de los estudiantes que desertaron o se trasladaron durante el segundo trimestre y 11 de 80 representa 13,8% que corresponde a los que desertaron o se trasladaron durante el tercer trimestre. Por ello, sin tomar en cuenta lo ocurrido en el primer trimestre, pues no se cuenta con la información de la matrícula, 37,5% de los estudiantes de octavo año abandonó el Colegio Hipotético en el segundo y tercer trimestre en el 2001.

2) Respecto al rendimiento académico se puede observar que no son muchos los estudiantes que tienen más de un 70 en cualquiera de los trimestres. Puede notarse que los datos para estas dos clases “de 70 a menos de 85” y “de 85 a 100” son muy similares en los tres trimestres, por lo que se podría pensar que son los mismos estudiantes los que tienen las mejores calificaciones y el resto termina por desertar o trasladarse de institución.

Desde un punto de vista general, si la información del cuadro fuera real, se tendrían resultados preocupantes en relación con la situación de los estudiantes de octavo año, pues se refleja bajos índices de promoción que pareciera estar provocando altas cifras de abandono. Esto amerita buscar alternativas para mejorar el rendimiento, con lo cual se espera que los jóvenes puedan permanecer todo el año en la institución y mejorar su estabilidad educativa.

A manera de recomendación didáctica para tratar este tema, el profesor debe preparar diferentes situaciones didácticas para que el estudiante pueda interactuar con los datos y construir cuadros que sean coherentes con la naturaleza de la información que se ha recolectado y con los intereses del estudio. Pero además, debe plantear representaciones tabulares para que los estudiantes puedan analizar en detalle la información que suministran las mismas.

## Representaciones gráficas

Las representaciones gráficas son un complemento importante de las representaciones tabulares. Los cuadros tienen el potencial de resumir mucho más información que una gráfica; sin embargo, desde un punto de vista visual, las gráficas permiten observar mejor el patrón de variación de algunos datos. Hay que recordar que para personas que no son especialistas en el área, el análisis visual se convierte en una herramienta fundamental para favorecer una interpretación adecuada de la información y observar sus patrones y tendencias.

La elección de una representación tabular o una gráfica depende de la naturaleza de los datos y del mensaje que se desea llevar al lector. No obstante, en términos generales, no es adecuado incluir dos representaciones, una tabular y una gráfica, para un mismo grupo de datos.

En general, las representaciones gráficas, junto a los cuadros, constituyen uno de los medios más difundidos para el resumen y presentación de la información. Permiten obtener una visión más clara y rápida acerca de lo que exhiben los datos. Es un instrumento que tiene por objeto presentar datos numéricos por medio de magnitudes geométricas, es decir, mediante longitudes, áreas, volúmenes, etc.

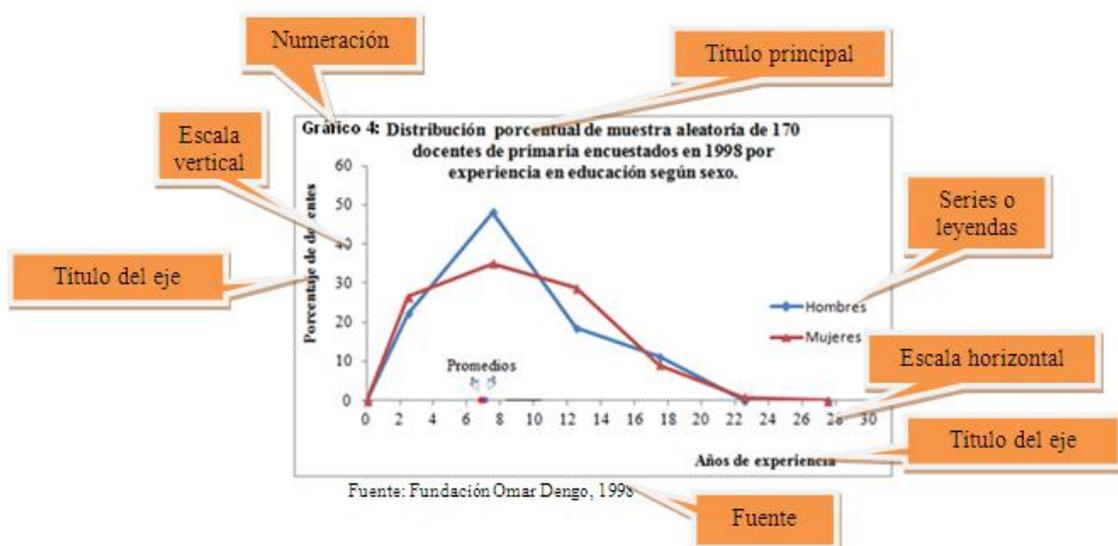
Las representaciones gráficas tienen algunos inconvenientes; entre ellos: solamente se puede presentar a la vez una cantidad limitada de datos y se dan valores aproximados.

Características que deben contener las representaciones gráficas:

- Proporciones adecuadas, ni muy largo ni excesivamente ancho.
- Explicarse por sí mismo (numeración, título, leyendas, símbolos, escalas y fuente).

- No se deben incluir muchas series de datos.
- Escalas bien definidas (si la escala se corta debe ser indicado).
- Debe ser sencilla, cómoda de interpretar y adecuada al tipo de información que se tiene.
- Los colores y el formato deben ser adecuados. No deben aparecer elementos innecesarios que distraigan la atención del lector.

A continuación se presenta una gráfica con los elementos necesarios:



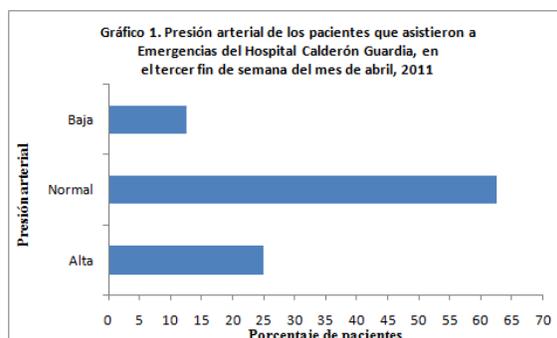
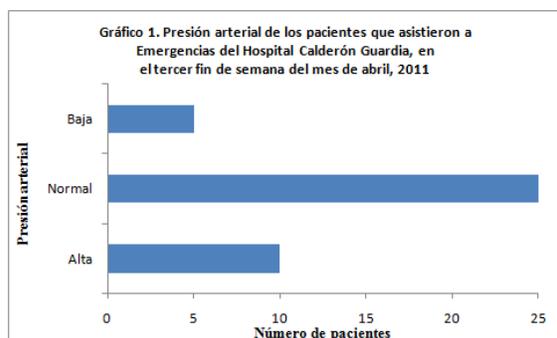
Existen muchas formas de representación gráfica para un análisis estadístico, las más comunes son las gráficas de barras presentes en diferentes medios comunicativos.

## Actividad 7

Con la información dada en la Actividad 3 de este apartado (*Representaciones*) proceda a buscar algún tipo de representación gráfica que le permita mostrar el estado de la presión arterial de los pacientes que asistieron a Emergencias del Hospital Calderón Guardia, en el tercer fin de semana del mes de abril del 2011.

## Análisis de la Actividad 7

Para realizar esta representación primeramente se requieren determinar las frecuencias de los distintos estados de la presión arterial que se registraron al momento de la consulta médica. Hay que recordar que la variable presión arterial se categorizó de la siguiente manera: *B significa baja*, *N normal* y *A alta*. Después de realizar el conteo correspondiente los resultados son: cinco con presión baja, 25 con la presión normal y 10 con la presión alta, los porcentajes correspondientes son respectivamente 12,5%, 62,5% y 25,0%. La forma más simple de representación grafica es la siguiente:



Cualquiera de estas dos gráficas puede ser utilizada para resumir esta información, como puede notarse ambas presentan un mismo patrón; aunque muchas veces la gráfica porcentual resulta más ilustrativa al momento de interpretar. Se observa que una importante mayoría (más de 60%) de los pacientes no tenía problemas con la presión arterial, pero resalta el hecho que 25% tenía problemas de hipertensión o presión alta, la cual es un grave problema de salud que debe ser controlado.

## Gráfica de barras simples

La gráfica que se ha utilizado se denomina de barras simples. Pueden ser barras horizontales como en este caso o barras verticales, se prefiere utilizar barras horizontales cuando la leyenda de las categorías corresponde a texto y barras verticales cuando dichas leyendas corresponden a valores numéricos. Esto se realiza con la intención de favorecer la lectura de las leyendas. Realice la siguiente actividad.

### Actividad 8

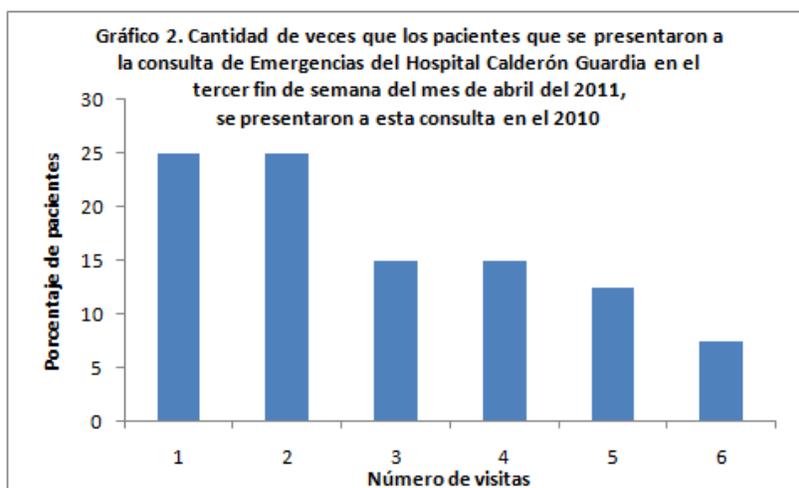
Con base en la información que se reportó en el cuadro 2 de este apartado, y que se presenta nuevamente, construya una gráfica de barras que represente la información.

**Cuadro 2. Cantidad de veces que los pacientes que se presentaron a la consulta de Emergencias del Hospital Calderón Guardia en el tercer fin de semana del mes de abril del 2011, visitaron dicha consulta en el 2010**

Número de veces	Cantidad de pacientes	Porcentaje de Pacientes
1	10	25,0
2	10	25,0
3	6	15,0
4	6	15,0
5	5	12,5
6	3	7,5
<b>Total</b>	<b>40</b>	<b>100</b>

## Análisis de la Actividad 8

Seguidamente se muestra una gráfica de barras verticales con los valores porcentuales.



Puede notarse como todos los pacientes atendidos en el tercer fin de semana del mes de abril del 2011, también debieron asistir a la consulta de Emergencias de este hospital en al menos una vez durante el 2010. La mitad de ellos había asistido una o dos veces, mientras que la otra mitad asistió entre tres y seis veces, inclusive, en dicho año.

\*\*\*

Desde el punto de vista de construcción de la gráfica, debido a que los valores de las categorías son números, se recurre a emplear barras verticales.

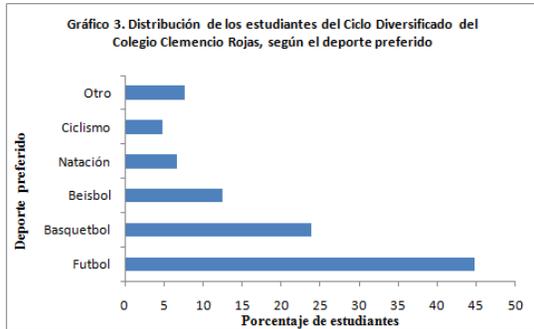
Otro aspecto a tomar en cuenta en las gráficas de barras es que si se utilizan variables cuantitativas o cualitativas ordinales, las barras se colocan siguiendo un orden natural tal como se muestra en las gráficas anteriores. No obstante, si la variable es cualitativa se recurre a ordenar las barras según la frecuencia, puede ser de mayor a menor o de menor a mayor según corresponda.

## Actividad 9

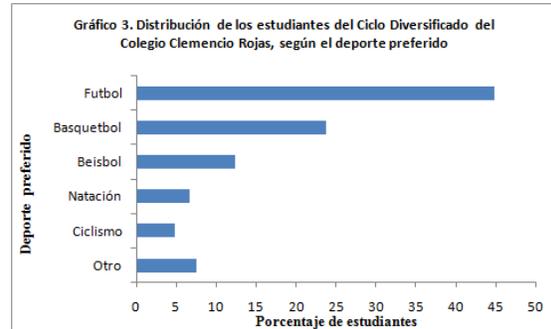
*Con la información dada en la Actividad 5 de este apartado, relacionada con la preferencia de los estudiantes del Ciclo Diversificado del Colegio Dominicó, construya un gráfico de barras simple.*

## Análisis de la Actividad 9

Debido a que la variable deporte preferido por los estudiantes es cualitativa nominal, cualquiera de las siguientes dos gráficas resulta adecuada.



Fuente: Encuesta realizada por el profesor de Educación Física del Colegio Dominico.



Fuente: Encuesta realizada por el profesor de Educación Física del Colegio Dominico.

Con estas gráficas se observa muy bien el patrón de variación y la preferencia por el deporte de los estudiantes de este colegio, donde el fútbol es el deporte preferido le sigue el básquetbol y el menos elegido fue el ciclismo.

**Nota:** Observe que la categoría de otro se incluye al inicio o al final, esto pues esa no responde al nombre de un deporte sino a la agrupación de todos los otros deportes citados por los estudiantes, pero que tuvieron una frecuencia muy baja.

## Gráficas circulares

Las gráficas circulares son utilizadas para comunicar la misma información que las de barras simples, con el agravante que su construcción es más compleja. También se les llama *de pastel o sectores*. Observe la siguiente gráfica:



Fuente: Colegio de Médico y Cirujanos de Costa Rica.

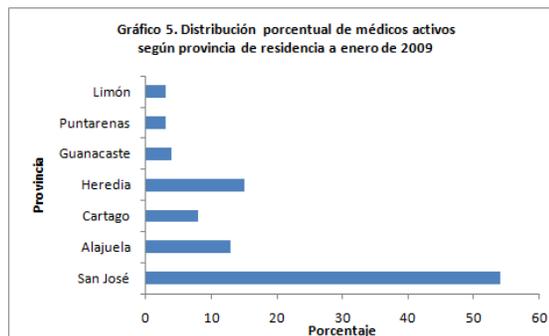
[www.medicos.sa.cr](http://www.medicos.sa.cr)

Aunque visualmente la gráfica anterior puede ser considerada más estética que una de barras, desde el punto de vista de la información que se quiere comunicar, fue necesario incluir los porcentajes para mostrar las diferencias entre las categorías, esto pues visualmente puede resultar complejo para un lector ver que sección es mayor, la

de Heredia o la de Alajuela. Esto pudo simplificarse presentando esta información en un cuadro, o en una gráfica de barras como se muestra:

**Cuadro 7. Distribución porcentual de médicos activos según provincia de residencia a enero de 2009**

Provincia de Residencia	Porcentaje
San José	54
Alajuela	13
Cartago	8
Heredia	15
Guanacaste	4
Puntarenas	3
Limón	3



Fuente: Colegio de Médico y Cirujanos de Costa Rica. [www.medicos.sa.cr](http://www.medicos.sa.cr)

Con cualquiera de las representaciones anteriores se comunica la misma información que en la gráfica circular previa. Por ello, se debe tener claro en qué momentos conviene utilizar una gráfica circular y en qué momentos puede ser sustituido por una técnica más simple.

Desde el punto de vista académico, se recomienda no abusar de este tipo de gráficas, aunque es adecuado analizar algunas de ellas, sobre todo para que el estudiante comprenda la información que suministra y pueda realizar interpretaciones de las mismas. Sin embargo, no se recomienda entrar en procesos de construcción de éstas, a no ser que se cuente con un recurso informático durante las lecciones; preferiblemente una hoja de cálculo o un programa para graficar datos estadísticos.

## Gráficas de barras comparativas

Las gráficas de barras tienen la virtud que permiten hacer comparaciones entre variables de modo que es posible visualizar relaciones entre ellas. Para una mayor ilustración desarrolle la siguiente actividad con una información que va a ser analizada con mayor detalle en la sección de probabilidades.

### Actividad 10

*La siguiente información corresponde a una muestra simulada de la relación existente entre el fumado por parte de las madres durante el embarazo y el bajo peso al nacer de sus hijos. Aunque los datos han sido simulados con fines académicos obedecen a un patrón que se ha determinado en varios estudios.*

**Cuadro 8. Distribución de la relación entre fumar durante el embarazo y el bajo peso al nacer en los niños para una muestra de 1000 partos**

Madres	Bajo peso al nacer		Total
	Sí	No	
Fumadoras	46	307	353
No fumadoras	39	608	647
Total	85	915	1 000

Utilice esta información para construir una gráfica que demuestre si existe relación entre el fumado durante el embarazo y el bajo peso del niño al nacer.

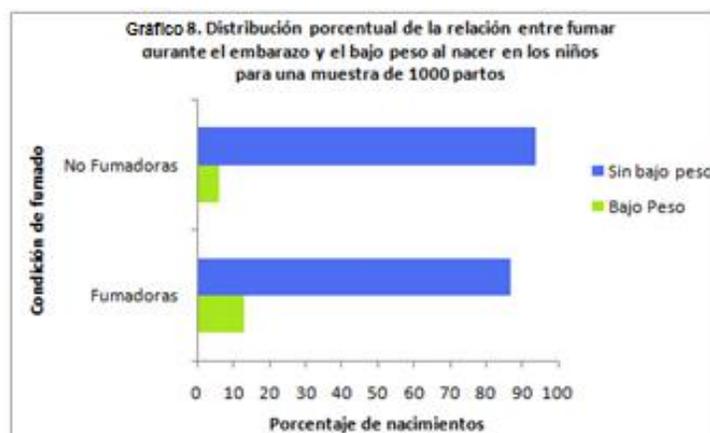
## Análisis de la Actividad 10

Con la información del cuadro es muy difícil poder identificar algún tipo de relación que permita establecer dicha relación. Para una mejor comprensión del fenómeno se requiere utilizar valores porcentuales, pero se debe tener cuidado pues como lo que se desea comparar es la condición de niños de bajo peso entre fumadoras y no fumadoras, los porcentajes deben ser calculados en esos dos grupos tal como se muestra en el cuadro siguiente:

**Cuadro 8. Distribución porcentual de la relación entre fumar durante el embarazo y el bajo peso al nacer en los niños para una muestra de 1000 partos**

Madres	Bajo peso al nacer		Total
	Sí	No	
Fumadoras	13,0	87,0	100
No fumadoras	6,0	94,0	100
Total	8,5	91,5	100

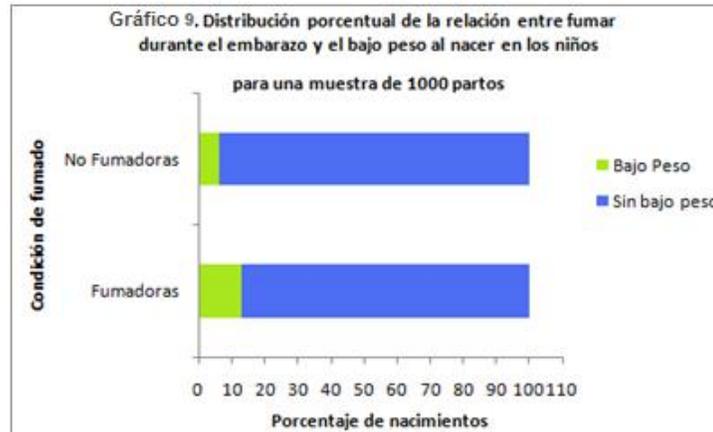
Con esta información es posible entonces construir la siguiente gráfica:



Observe que aunque los comportamientos para fumadoras y no fumadoras parecieran similares, en realidad el porcentaje de nacimientos de niños de bajo peso es más del doble entre las fumadoras y las no fumadoras, lo que refleja que, al menos para esta muestra de 1000 partos, efectivamente se observa una relación.

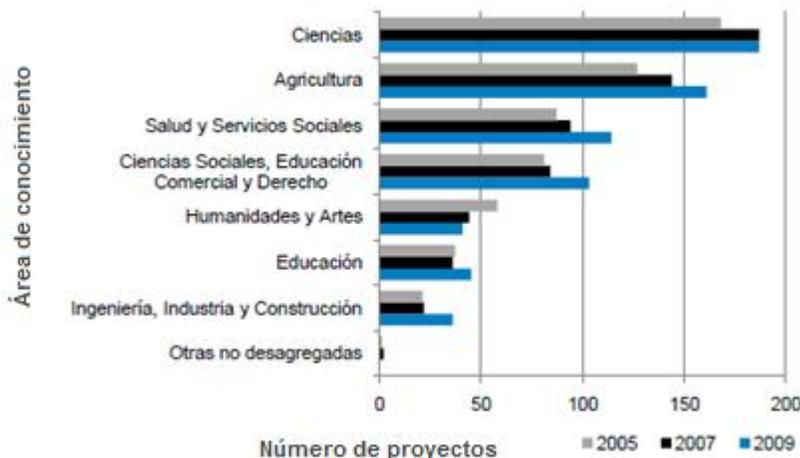
## Gráficas comparativas al 100%

Retomando la información suministrada en la Actividad 10, otra forma de realizar la comparación entre condición de fumado de la madre y el bajo peso del niño al nacer es por medio de una gráfica como la siguiente:



A estas representaciones se les llama, **gráficas de barras al 100%** permiten realizar análisis comparativos similares a los de las barras comparativas. Esta es una buena representación cuando hay pocas categorías para cada subcategoría como en el caso anterior, pero el análisis visual se hace complejo cuando este número es muy grande y es preferible recurrir a la gráfica de barras comparativas. Observe el siguiente ejemplo.

Gráfico 10. Proyectos de investigación en institutos, centros, laboratorios y estaciones experimentales de la UCR, según área del conocimiento<sup>a/</sup>



a/ Grupos amplios de la Clasificación Internacional Normalizada de la Educación (Unesco, 1997).

En esta gráfica existen muchas categorías para la variable *Área del conocimiento*, para cada categoría se analizan tres años, por lo que la anterior es la mejor estrategia de presentación para esos datos.

## Gráficas de líneas

En algunos casos las gráficas de barras pueden ser reemplazadas para analizar de una mejor manera el patrón de comportamiento o tendencia de un conjunto de datos. Este es el caso en que los datos representan series de tiempo, en los cuales la tendencia lineal es clara. Realice la siguiente actividad.

### Actividad 11

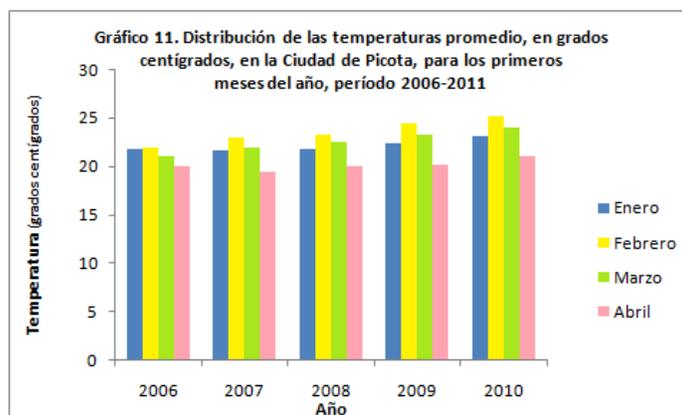
Los siguientes datos corresponden a la temperatura promedio por mes en cierta ciudad para el período que va del 2006 al 2010. Utilice esta información para representar gráficamente este comportamiento.

**Cuadro 9: Distribución de las temperaturas promedio, en grados centígrados, en la Ciudad de Picota, para los primeros meses del año, período 2006-2010**

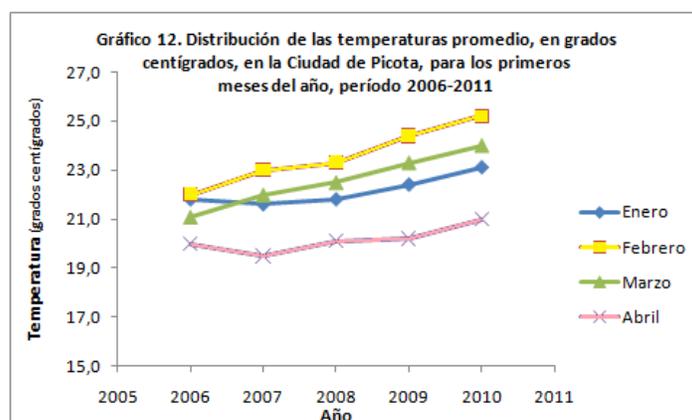
Mes	Año				
	2006	2007	2008	2009	2010
Enero	21,8	21,6	21,8	22,4	23,1
Febrero	22,0	23,0	23,3	24,4	25,2
Marzo	21,1	22,0	22,5	23,3	24,0
Abril	20,0	19,5	20,1	20,2	21,0

### Análisis de la Actividad 11

Si se utiliza una gráfica de barras comparativas, se obtiene lo siguiente:



Aunque se puede realizar un análisis en función de la variación que se presenta en las temperaturas año tras año, no es ésta la mejor forma de comunicar el mensaje al lector, en estos casos se recomienda la siguiente gráfica.



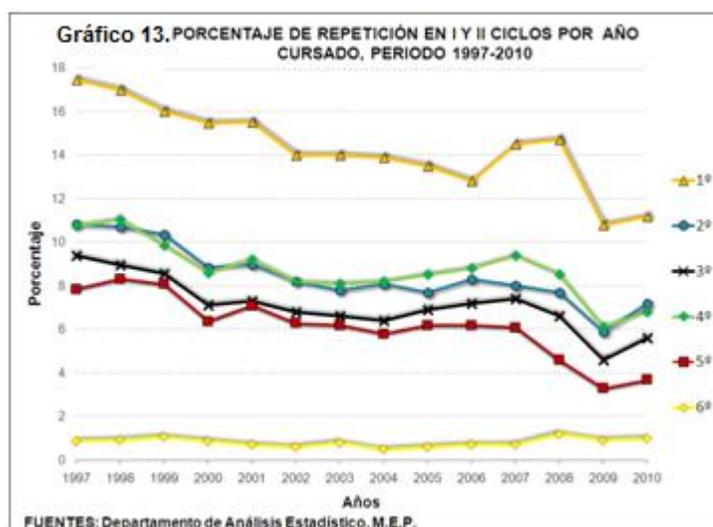
Puede notarse que ahora el patrón es claro, se observa un patrón creciente en la temperatura promedio mensual para los cuatros meses en este período, además el mes de abril muestra temperaturas relativamente más bajas que los otros meses, así como un crecimiento de temperatura más lento, mientras que el mes con mayores temperaturas es febrero.

Como se indicó antes, **las gráficas lineales o de línea** se convierten en una importante herramienta para resumir datos que representan series de tiempo (los datos son clasificados de acuerdo con los meses, años, periodos, etc.), tal como se observó en el ejemplo anterior. A continuación se presenta otro caso:

\*\*\*

## Actividad 12

Analice la información de la siguiente gráfica y responda las interrogantes que se le plantean.



a) ¿Cuál es la importancia de recabar información estadística en relación con el tema educativo?

b) ¿Cómo se debe interpretar el concepto de porcentaje de repetición, que está presente en la gráfica?

- c) Al comparar la repetición entre los diferentes niveles escolares ¿Qué relaciones encuentra?
- d) ¿Qué diferencias se presentan en los porcentajes de reprobación de sexto año respecto a los de primer año?
- e) ¿En qué años los porcentajes de repetición de primer año superaron 14%?
- f) En términos generales ¿Qué evolución tuvo la repetición de los niveles escolares entre el 2000 y el 2010?

## Análisis de la Actividad 12

- a) La importancia de recabar este tipo de información estadística es brindarles a las autoridades del MEP, a los usuarios de las estadísticas, a los investigadores y a otros entes relacionados con la educación, datos acerca del comportamiento del sistema educativo en relación con ciertas variables de interés. Esto tiene el propósito de analizar los resultados obtenidos en estas investigaciones y así tomar decisiones trascendentales para el país.
- b) El porcentaje de repetición que está presente en la gráfica se interpreta como el número de estudiantes que están repitiendo por cada 100 estudiantes matriculados. Por ejemplo, en primer grado del 2010 por cada 100 estudiantes matriculados 11 eran repitentes y en sexto grado por cada 100 estudiantes matriculados uno era repitente.
- c) En primer lugar, se observa la marcada diferencia entre tres escenarios, dos de ellos extremos, la alta repetición en primer grado y la baja repetición en sexto grado, y un escenario central donde están los otros grados, y muy lejos de estos extremos. Además, puede notarse que segundo y cuarto tienen un comportamiento muy parecido, mientras que tercero y quinto que tenían un comportamiento muy parecido entre 1997 y 2005, después de eso tienden a separarse.
- d) Se puede analizar que la diferencia de los porcentajes de repitencia de sexto año respecto a los de primer año al inicio del período (1997) era alrededor de 16 puntos porcentuales, mientras que al final del periodo (2010) es de aproximadamente 10 puntos porcentuales. Es importante indicar que en estos casos no se habla de una diferencia en términos de porcentajes sino de puntos porcentuales, pues los conceptos son diferentes. Por ejemplo, si se habla de una diferencia de 16% significa que sobre la base del primer valor el segundo es 16% más grande, mientras que si habla de 16 puntos porcentuales significa que la diferencia entre los dos porcentajes es de 16 puntos.
- e) Los años en que los porcentajes de repetición de primer año superaron 14% son de 1997 al 2001, 2007 y 2008.  
*Como una actividad adicional se le puede pedir a los estudiantes que conjeturen sobre la repitencia en primer grado, tomando en cuenta la información de la gráfica.*
- f) Puede notarse que entre el 2000 y el 2005 se mantienen muy estables, aumentan entre 2006 y 2007, presentando un importante descenso entre 2008 y 2009, con un aumento nuevamente en 2010. No obstante, los grados de primero y sexto,

experimentan patrones un poco diferentes. Por ejemplo, en sexto grado el porcentaje de repetición se mantuvo relativamente constante en todo el periodo; en cambio en primer grado el comportamiento fue muy irregular ya que entre 1997 y 2002 tendió a bajar, luego se mantuvo entre 2002 y 2004, seguidamente entre el 2004 y 2006 descendió, entre 2006 y 2008 creció para luego bajar considerablemente en el 2009 y volver a crecer en el 2010.

Análisis como el planteado anteriormente, permite que los estudiantes puedan valorar la importancia de recolectar y publicar información estadística, como una estrategia para diferentes actividades, aquí se utiliza como un instrumento para analizar el comportamiento del sistema educativo. Asimismo, es importante que los jóvenes estén en capacidad de realizar una lectura lo más completa posible, así como que utilicen la información para conjeturar y argumentar hipótesis utilizando los datos proporcionados.

## Diagramas de puntos

Cuando se tiene un conjunto de datos correspondiente a una variable cuantitativa, se tiene la desventaja que los gráficos anteriores no permiten resumir fácilmente la información de modo que pueda representarlos por algún tipo de gráficos de barras. Para llevar a cabo un análisis con este tipo de información se debe buscar algunas estrategias para analizar, desde un punto de vista exploratorio, el comportamiento de esos datos.

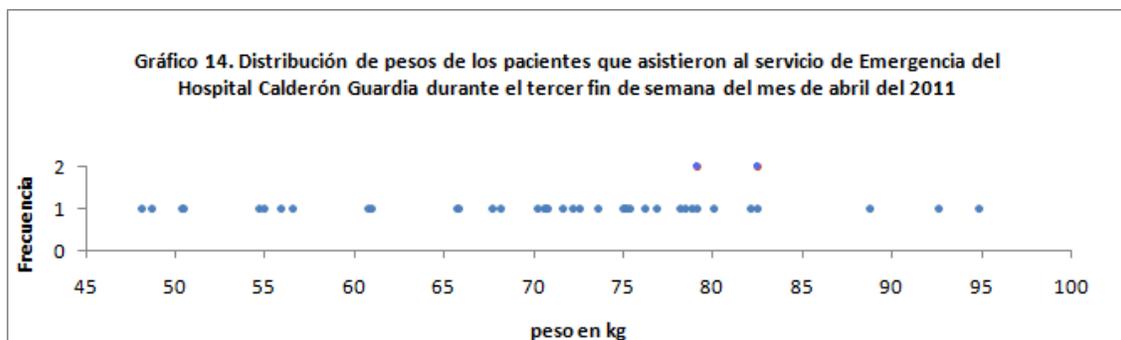
### Actividad 13

*Con la información dada en la Actividad 3 de este apartado, proceda a buscar algún tipo de representación gráfica que le permita representar el peso de los pacientes que asistieron al servicio de Emergencias del Hospital Calderón guardia en el tercer fin de semana del mes de abril del 2011.*

*Con la información de esta variable, determine si existen importantes diferencias entre el peso de los pacientes por sexo.*

### Análisis de la actividad 13

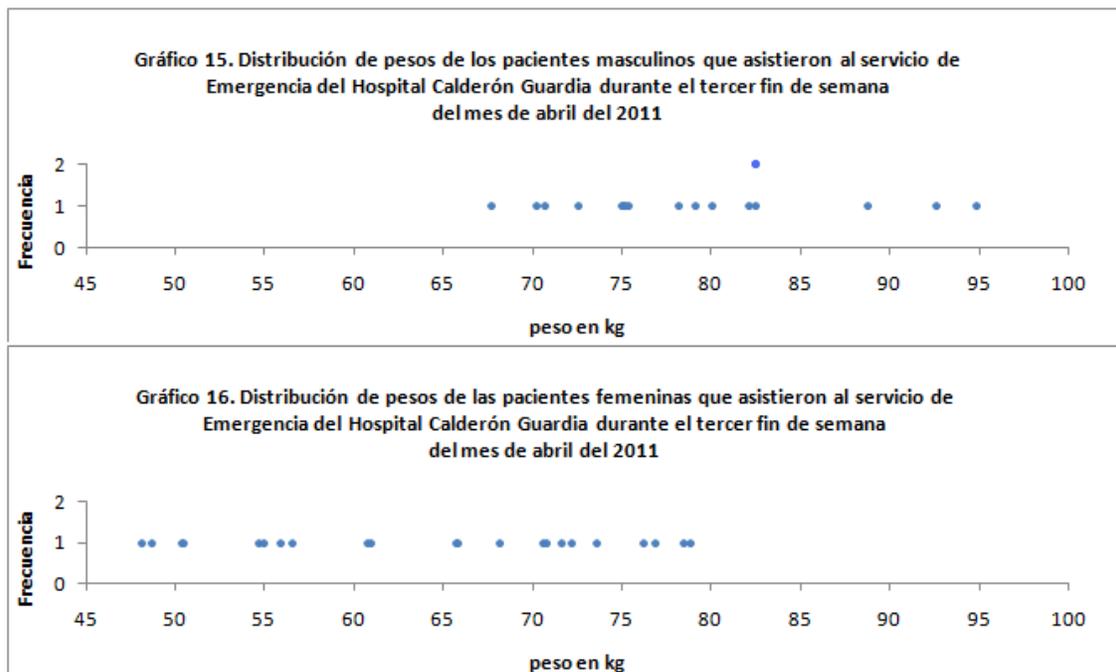
En este caso se tienen 40 datos muy diferentes unos de otros, por lo que una representación gráfica de barras o circular no es adecuada pues la cantidad de barras o sectores serían demasiados todos con muy poca frecuencia. En este sentido una forma de representación es la siguiente:



Este tipo de gráfica permite evidenciar la distribución de los datos de modo que pueden observarse los pesos donde hay más concentración en este caso entre los 65 kg y los 83 kg aproximadamente, también se observa que tan variable es la información, en este caso se nota gran variabilidad pues hay mucho espacio entre los puntos, de hecho ellos tienen un rango de variación de aproximadamente 50 kg (entre 48 kg y 95 kg).

Aunque no se acostumbra emplear estas gráficas para presentarlas a lectores, son un valioso recurso para analizar en forma exploratoria los datos y comprender mejor su comportamiento antes de realizar otros análisis estadísticos.

También se utilizan para efectuar comparaciones entre variables, esto puede observarse en las siguientes gráficas.



Al comparar las dos gráficas claramente se observa que los pesos de los varones son, en general, más altos que los de las mujeres. El peso de los varones tiene fuerte concentración entre 70 kg y los 83 kg; mientras que el peso de las mujeres presenta una concentración muy heterogénea, tomando valores entre 48 kg y 79 kg, con una ligera agrupación entre 70 kg y 79 kg. Por esta misma razón, los pesos de las mujeres presentan más variabilidad.

El análisis de la posición y la variabilidad se aborda con mayor detalle en la sección de medidas estadísticas.

## Histogramas y polígonos de frecuencias

Estos son tipos muy particulares de gráficas, al igual que los diagramas de puntos se utilizan para representar información de variables cuantitativas continuas, solo que ahora se requiere que dicha información haya sido resumida previamente en una distribución de frecuencias. Analice detenidamente la siguiente situación y trate de dar una representación adecuada a la información que se le brinda.

## Actividad 14

Se requiere volver a analizar el peso de los pacientes que asistieron al servicio de Emergencias al Hospital Calderón Guardia. Con dicha información realice las siguientes actividades.

a) Construya una distribución de frecuencias para los pesos de los hombres y otra para el peso de las mujeres. En el caso de los hombres utilice las siguientes clases o grupos: de 65 a menos de 70, de 70 a menos de 75, ..., de 90 a menos de 95. Por otro lado, para el caso de las mujeres las clases serán: de 45 a menos de 50, de 50 a menos de 55, ..., de 75 a menos de 80.

b) Con la información de las distribuciones de frecuencia utilice algún tipo de representación gráfica para comparar el peso entre hombres y mujeres.

## Análisis de la Actividad 14

a) Con base en los datos y en las clases que se sugirieron para la actividad, los cuadros que resumen la información son las siguientes:

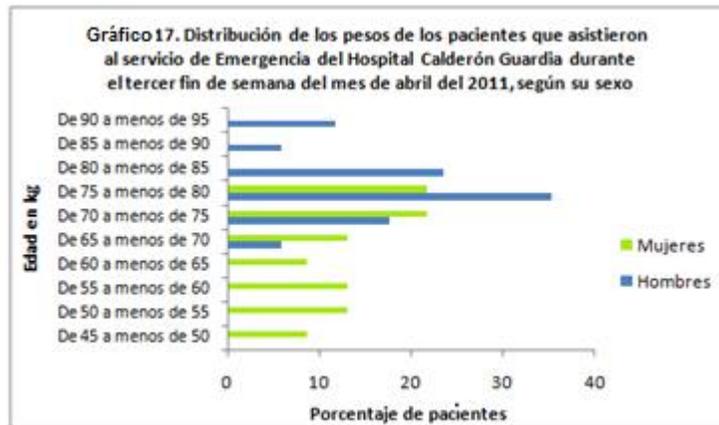
**Cuadro 10. Distribución de los pesos de las pacientes femeninas que asistieron al servicio de Emergencia del Hospital Calderón Guardia durante el tercer fin de semana del mes de abril del 2011**

Peso de las Pacientes (kg)	Cantidad de pacientes	Porcentaje de pacientes
De 65 a menos de 70	1	5,9
De 70 a menos de 75	3	17,6
De 75 a menos de 80	6	35,3
De 80 a menos de 85	4	23,5
De 85 a menos de 90	1	5,9
De 90 a menos de 95	2	11,8
<b>Total</b>	<b>17</b>	<b>100,0</b>

**Cuadro 11. Distribución de los pesos de los pacientes masculinos que asistieron al servicio de Emergencia del Hospital Calderón Guardia durante el tercer fin de semana del mes de abril del 2011**

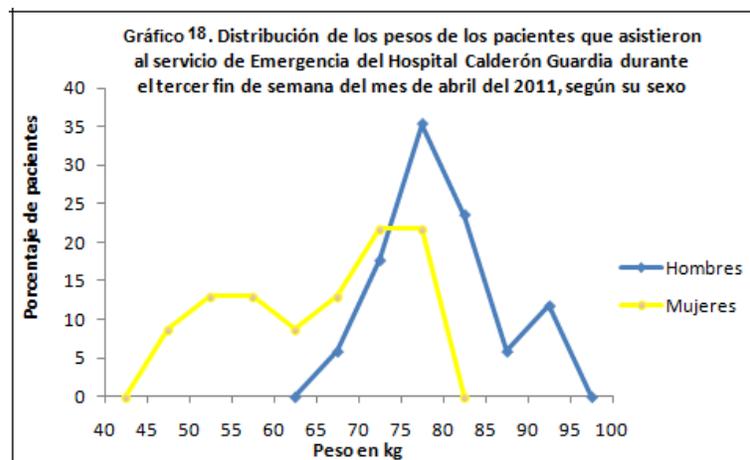
Peso de los Pacientes (kg)	Cantidad de pacientes	Porcentaje de pacientes
De 45 a menos de 50	2	8,7
De 50 a menos de 55	3	13,0
De 55 a menos de 60	3	13,0
De 60 a menos de 65	2	8,7
De 65 a menos de 70	3	13,0
De 70 a menos de 75	5	21,7
De 75 a menos de 80	5	21,7
<b>Total</b>	<b>23</b>	<b>100,0</b>

b) Se puede pensar en diferentes estrategias gráficas para representar estos datos, entre ellos una gráfica de barras comparativas tal como la siguiente.



Con esta gráfica se puede realizar un análisis similar al que se efectuó con los diagramas de puntos; pero se puede hacer muy confuso para la interpretación de un lector. Por esta razón, se recurre a utilizar otro tipo de representación gráfica que ayude con la interpretación.

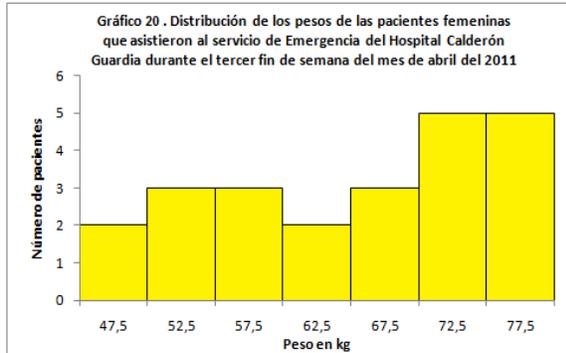
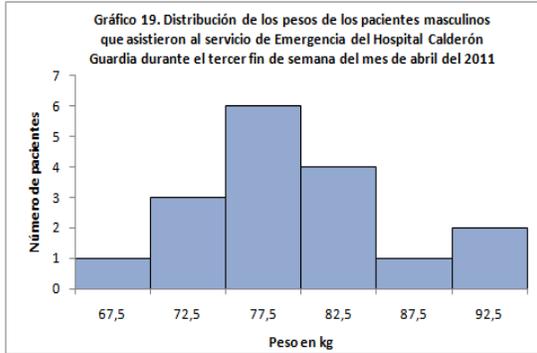
Uno de ellos es el **polígono de frecuencias**, que consiste en un tipo particular de gráfica lineal que se construye tomando como abscisa el punto medio de la clase y como ordenada la frecuencia absoluta o porcentual de la clase, se cierra al principio y al final. A continuación se presentan los polígonos de frecuencias para los pesos de los pacientes.



Como puede notarse esta es una gráfica muy ilustrativa respecto al comportamiento de los pesos de los pacientes. Los pesos de los hombres son superiores a los de las mujeres, la mayor concentración de pesos de los varones están entre 68 kg y 83 kg; mientras que el peso de las mujeres tiene una ligera concentración entre 67 kg y 80 kg. Además, se observa también mayor variabilidad en los pesos de las mujeres.

Otra representación gráfica que es muy utilizada cuando los datos se encuentran resumidos en una distribución de frecuencias, es el **histograma**, que consiste en una gráfica de barras verticales, donde las barras no guardan separación entre sí.

En el eje de las abscisas se utilizan los que representan los puntos medios de las clases o los límites de cada clase, la altura de la barra es la frecuencia absoluta o relativa. Al igual que en otras gráficas la escala no necesita empezar en cero, las frecuencias se presentan en la escala vertical, la cual debe empezar de cero. Seguidamente se presentan los histogramas correspondientes a los pesos de los pacientes.

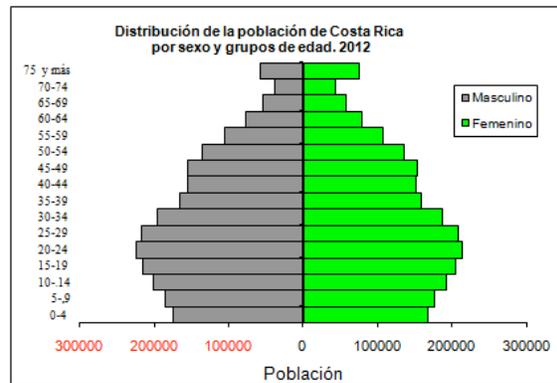
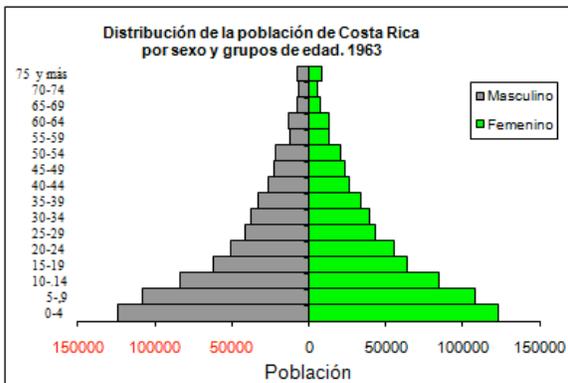


\*\*\*

Para finalizar esta sección es importante hacer notar que existen muchas otras representaciones gráficas que pueden ser utilizadas, por lo que el docente debe plantear diferentes situaciones que le ayuden al estudiante a adquirir las habilidades necesarias para leer e interpretar la información gráfica. La siguiente actividad es un ejemplo particular de estas situaciones.

### Actividad 15

*Un adagio popular indica “una imagen es mejor que mil palabras”, seguidamente se presentan dos gráficas denominadas pirámides de población, en este caso se utilizan para representar la distribución de la población de Costa Rica en los años 1963 y una proyección al 2012.*



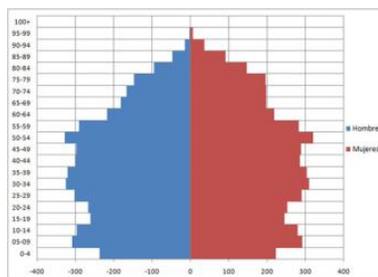
Fuente: Centro Centroamericano de Población, [ccp.ucr.ac.cr](http://ccp.ucr.ac.cr)

- ¿Cuáles son las principales diferencias que se observa entre las distribuciones de las dos poblaciones? ¿cómo se pueden interpretar esas diferencias?
- ¿En cuál de los años el porcentaje de población adulta mayor (65 o más años) es mayor?

- c) ¿En cuál de estos años existe un mayor porcentaje de la población en edad escolar y colegial?
- d) ¿Por qué es necesario utilizar porcentajes para realizar las comparaciones anteriores?
- e) En el grupo de edad entre 0 y 4 años, ¿qué sexo tiene más población?

### Análisis de la Actividad 15

- a) En primer lugar se debe observar que la forma de las gráficas es diferente. En 1963 tiene forma de una pirámide casi perfecta mientras que para el 2012 la base se ha estrechado ¿qué significa eso? Debido a que, en términos relativos, la base de la gráfica es mucho más amplia en 1963 que en el 2012, eso significa que la proporción de población de menores de 15 años era mucho mayor en 1963 que en el 2012. De igual forma, en términos relativos, al ver la cúspide de las pirámides, en los grupos de edades de adultos mayores las barras del 2012 son más largas que las de 1963, ello significa que la proporción de población adulta mayor es superior en el 2012 respecto a 1963. Por todo esto, se concluye que la estructura de la población del país en el 2012 es más vieja que la de 1963, es decir en el 2012 hay más porcentajes de gente adulta y adulta mayor que en 1963, donde la población era mucho más joven. Este es un fenómeno típico que muestra una disminución en los nacimientos, lo cual se presenta en mayor medida en países que han tenido una importante evolución demográfica, como los países europeos, la siguiente pirámide es un ejemplo de la forma en que se distribuye la población en países de alta evolución demográfica.



- b) En segundo lugar se deben observar diferencias por sexo, aunque en la primera barra hay más hombres que mujeres, producto de que se presentan mayor nacimiento de varones, en edades adultas la cantidad de mujeres supera al de los hombres, lo que indica que los hombres se mueren a edades más tempranas que las mujeres. Este patrón, es similar para ambos años, aunque es más notorio para el 2012. Estas son conjeturas realmente importantes y no se pueden pasar por alto en un análisis de este tipo, aunque podría ser que se requieran nuevas preguntas para explotar la información, esta es la tarea del docente.
- c) También se observa que entre 1963 y el 2012 se ha presentado en el país un fuerte crecimiento poblacional, para todos los grupos de edades la cantidad de población es mucho mayor en este último año.
- d) Además, también se observa que el promedio de vida es más alto en el 2012, mayores porcentajes de población viven más de 60 años.

## V. Medidas resumen de posición y de variabilidad

### Introducción

Como se viene ilustrando en los apartados anteriores, una vez recolectados los datos se requiere establecer estrategias que ayuden a resumirlos, para extraer información relevante, especialmente cuando las variables son cuantitativas. Por medio de métodos gráficos, se puede tener una representación visual de los tres aspectos fundamentales en un conjunto de datos para una determinada variable, a saber: la forma de su distribución, su posición o localización; es decir, alrededor de qué valor se tienden a concentrar en mayor medida los datos y la dispersión o variabilidad de las observaciones alrededor de los valores centrales. Sin embargo, en muchas ocasiones se requiere establecer valores numéricos que ayuden a caracterizar esos datos.

En esta sección se introducen algunas medidas estadísticas que permiten analizar la posición y la variabilidad de los datos.

### Actividad 1

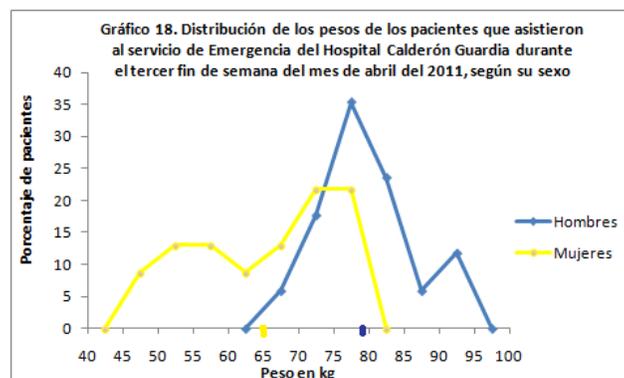
*Seguidamente se retoma un análisis planteado en la sección anterior, con la intención de complementar el estudio sobre los pacientes que asistieron al servicio de Emergencias del Hospital Calderón Guardia en el tercer fin de semana del mes de abril del 2011. Hay que recordar que el propósito básico es caracterizar ese grupo de pacientes en función de las variables del estudio.*

*De acuerdo con esos datos y con los análisis efectuados en la sección anterior, proceda a caracterizar el comportamiento de cada variable según lo que se solicite en cada caso.*

- 1) ¿Cuál es el sexo que asistió con más frecuencia al servicio de Emergencias de dicho hospital en las fechas investigada? ¿Cuál es el cantón en el que se presentó la mayor frecuencia de pacientes en ese fin de semana?*
- 2) ¿Cuál es el número más común de asistencias al servicio de emergencias durante el 2010 entre los pacientes que asistieron a este servicio en el tercer fin de semana del mes de abril del 2011?*
- 3) Aunque de los análisis realizados en la sección previa se determinó que las edades de los pacientes analizados se concentraban, en mayor medida, entre los 35 y los 55 años (ver cuadro 3 de la sección anterior), interesa ahora resumir esta información mediante un solo valor ¿Cuál sería ese valor?*
- 4) Del mismo modo, en el estudio previo (ver gráficos 15, 16 y 18 de la sección anterior) se determinó que los pesos de las mujeres son menores que los de los varones; no obstante, es importante poder resumir los datos de cada grupo en función de una medida que resuma adecuadamente todos los demás datos. Además, identificar los valores extremos de cada grupo.*

## Análisis de la Actividad 1

- 1) Debido a que llegaron 23 mujeres y 17 varones, se puede decir que el sexo más común entre estos pacientes es el femenino. Por otro lado, de acuerdo con el cuadro 1 (apartado anterior), se presentaron pacientes al servicio de Emergencias de seis cantones de la provincia San José, los cantones con más frecuencia de pacientes fueron San José y Moravia, ambos con nueve pacientes. Regularmente el dato con mayor frecuencia dentro de un grupo, se le denomina *moda o valor modal*. En el caso del sexo de los pacientes la moda es mujer, mientras que en el caso del cantón de residencia de los pacientes se presentaron dos modas, los cantones San José y Moravia.
- 2) En el cuadro 2 de la sección previa, se determinó que 10 de los pacientes que asistieron a Emergencias en el tercer fin de semana de abril habían asistido una vez durante el 2010 y otros 10 de ellos asistieron dos veces durante el 2010. Por esta razón, puede decirse que la variable número de visitas a Emergencias en el 2010, tiene dos *modas* que son una y dos veces.
- 3) Ante la necesidad de resumir los datos de la variable edad de los pacientes, se debe buscar una estrategia que permita aglutinar estos datos en un solo número. Un valor que es muy conocido se obtiene de sumar todas las edades y dividirlos entre el número total de pacientes, el cual se conoce como *edad promedio o promedio* de las edades. Después de realizar el procedimiento se obtiene que la edad promedio es 45,4 años. Este valor es muy representativo en cuanto a la posición de la distribución de los datos, ya que como puede notarse efectivamente se encuentra en el rango en que se concentra la mayor parte de las edades de los pacientes.
- 4) Aplicando el mismo procedimiento anterior se puede establecer que el peso *promedio* de los varones fue 78,8 kg, mientras que el peso *promedio* de las mujeres fue 64,8 kg. Si se analiza nuevamente el gráfico 18 de la sección anterior se tiene:



En el eje de las abscisas se han marcado los promedios para cada caso, nuevamente estos valores tienden a ubicarse en una posición en la cual existe una gran concentración de datos.

Por otro lado, los valores extremos de cada uno de los grupos de pesos son:

Hombres: el *mínimo* fue 67,7 kg y el *máximo* 94,9 kg.

Mujeres: el *mínimo* fue 48,1 kg y el *máximo* 79,1 kg.

Los resultados de la actividad anterior dejan entrever el uso de algunas medidas de resumen de la información, seguidamente se analizan algunas propiedades de cada una de ellas.

## Moda

Corresponde al dato que más se repite en un conjunto de datos, se utiliza tanto para datos cualitativos como cuantitativos. En un grupo de observaciones puede ser que no exista *moda* o que haya más de una *moda*. Se debe tener cuidado con su interpretación, debido a que el concepto de valor que más se repite o de mayor frecuencia, no significa que la mayoría de los datos tome ese valor, pues la mayoría representa más de la mitad, pero el valor modal no siempre incluye más de la mitad de las observaciones.

## Promedio o media aritmética

El *promedio* o *media aritmética* es uno de los valores más utilizados dentro de los análisis estadísticos, es la medida de *tendencia central* de mayor relevancia. Se dice que es una medida de tendencia central debido a que tiende a caracterizar la mayor concentración de datos, tal como se pudo evidenciar en la actividad previa. En general, si para una variable, las observaciones generan  $n$  datos que se pueden representar con  $X_1, X_2, \dots, X_n$  entonces el promedio se calcula mediante la fórmula:

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

Es decir, se suman todas las observaciones y se divide entre el número total de ellas.

## Máximo y mínimo

Estos valores se definen a partir de sus propios nombres. El *máximo* corresponde al dato de mayor valor numérico del conjunto y el *mínimo* representa el de menor valor numérico.

**Nota:** El promedio tiene la desventaja de ser muy sensible a los valores extremos, de modo que si un conjunto de datos tiene unos pocos valores más grandes o más pequeños del común, provocan que el promedio se sesgue o incline hacia esos valores extremos. Para evidenciar esto resuelva la siguiente actividad.

## Actividad 2

*En un curso con alto grado de dificultad de la universidad, un estudiante debió realizar cinco exámenes parciales y debía obtener una nota mínima de 7,0 para aprobar la materia. Desafortunadamente en el cuarto examen tuvo un problema de salud que afectó su concentración y le hizo obtener una baja calificación. Las notas obtenidas son las siguientes: 7,1; 7,5; 7,8; 2,8 y 8,0.*

*De acuerdo esta información determine:*

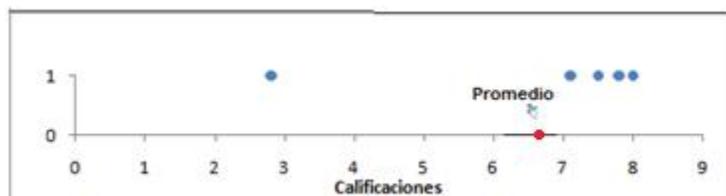
1. La nota promedio del estudiante en ese curso.
2. Considera usted justa esta calificación, si la respuesta es negativa; establezca algún procedimiento que permita otorgarle una calificación más justa al estudiante.

**Nota:** La situación planteada es hipotética, ha sido redactada con fines didácticos. Por esta razón, el análisis del problema se puede realizar desde una perspectiva eminentemente estadística, tomando como referente el aporte que las diferentes técnicas pueden dar para resolverlo.

## Análisis de la Actividad 2

- 1) La nota promedio de este estudiante está dada por  $\frac{7,1+7,5+7,8+2,8+8,0}{5} = 6,64$ , por lo que estaría perdiendo el curso y deberá realizar un examen de ampliación que es similar al examen de convocatoria que se realiza en el colegio.
- Evidentemente este resultado no es justo, pues, con excepción de la nota del cuarto examen, el estudiante fue superando sus calificaciones. El principal problema que se presenta obedece a la nota del cuarto examen. Para tratar de una manera más justa al estudiante, se podrían plantear varias posibilidades, entre ellas:
    - i. Eliminar la nota del cuarto examen y obtener el promedio con los cuatro exámenes restantes, el cual sería:  $\frac{7,1+7,5+7,8+8,0}{4} = 7,60$ , por lo que estaría aprobando el curso.
    - ii. Sustituir la nota del cuarto examen por el promedio de los otros cuatro, por lo que la nota obtenida sería la misma obtenida anteriormente  $\frac{7,1+7,5+7,8+7,6+8,0}{5} = 7,60$ , es decir estas dos alternativas son equivalentes.
    - iii. Eliminar la calificación más baja y la más alta, luego determinar el promedio de los tres exámenes restantes  $\frac{7,1+7,5+7,8}{3} = 7,47$ , por lo que nuevamente estaría aprobando el curso.
    - iv. Solicitar al estudiante que reponga el cuarto examen y con ese valor promediar nuevamente las calificaciones.

La situación planteada en esta actividad deja en evidencia el efecto que uno o más valores extremos ejercen sobre el promedio. Se puede notar que el valor 2,8 del cuarto examen provoca que el promedio se sesgue hacia un valor bajo, muy lejos del común de las notas. Esta situación se observa en el siguiente gráfico, en el cual se marca sobre el eje de las abscisas el valor del promedio:



## VI. Recomendaciones metodológicas

Como se desarrolló en la fundamentación teórica de los nuevos programas de estudio, se promueve el énfasis en una organización de las lecciones, con base en cuatro pasos o momentos centrales, a saber:

1. Propuesta de un problema para iniciar la lección.
2. Trabajo estudiantil independiente
3. Discusión interactiva y comunicativa.
4. Clausura o cierre.

Para ilustrar esta propuesta, se presenta la siguiente situación, relacionada con el desarrollo de algunas habilidades propuestas para séptimo año.

Conocimientos	Habilidades específicas
Recolección de información: experimentación e interrogación	Recolectar datos del entorno por medio de experimentación e interrogación.
Frecuencia: absoluta y porcentual	Utilizar los conceptos de frecuencia absoluta y porcentual para resumir un conjunto de datos.
Representaciones. Tabular: cuadros de frecuencia absoluta y porcentual.	Resumir un conjunto de datos por medio de cuadros simples y comparativos con frecuencia absoluta y porcentual.

Si se quiere desarrollar en los estudiantes estas habilidades, se deberían planear los siguientes cuatro momentos:

### Propuesta de un problema

Lo primero que un docente debería hacer es motivar a los estudiantes para que sientan interés de realizar la actividad planteada, en este caso se les puede decir que con la actividad a realizar, además de conocer cómo se hace la recolección de datos y su organización, van a conocer acerca de las preferencias de sus compañeros de clase, lo cual es importante en este nivel escolar ya que vienen cambiando de ciclo y de ambiente.

#### Problema

*El paso de la escuela al colegio trae consigo un sin número de emociones, algunas de ellas positivas y otras que pueden desencadenar temores, al desconocer las personas con las que se va a encontrar en su sección.*

*El profesor del grupo 7-1 conocedor de la preocupación de sus estudiantes, decide realizar una actividad donde; además, de desarrollar habilidades en Estadística, los*

estudiantes tengan la oportunidad de interactuar entre ellos y conocer sus gustos en diversos aspectos.

El profesor considera que algunos de los aspectos más importantes para conocerlos y caracterizarlos son: predilección en programas de televisión y la cantidad de horas que dedican a mirarla; tipo de comida y bebida que le gusta más y la frecuencia con la que injiere “comidas rápidas” a la semana, dicha frecuencia clasificada en muchas veces, algunas veces y nunca; deporte preferido y cantidad de horas a la semana que destina para practicar dicho deporte; materia preferida y cantidad de horas que destina a la semana para prepararse en matemática.

El profesor pide a sus estudiantes que hagan grupos de cuatro o cinco personas, para que cada subgrupo realice la caracterización de un aspecto. El profesor solamente les da el tema a cada subgrupo y espera antes de dirigirlos.

*Grupo 1: predilección en programas de televisión y la cantidad de horas que dedican a mirarla.*

*Grupo 2: tipo de comida y bebida que le gusta más y la frecuencia con la que injiere “comidas rápidas” a la semana, dicha frecuencia clasificada en muchas veces, algunas veces y nunca.*

*Grupo 3: deporte preferido y cantidad de horas a la semana que destina para practicar dicho deporte.*

*Grupo 4: materia preferida y cantidad de horas que destina a la semana para prepararse en matemática.*

## Trabajo estudiantil independiente

En esta etapa se espera que los estudiantes lean la situación planteada, diseñen una estrategia para la recolección de la información de todos sus compañeros de clase.

Los estudiantes pueden utilizar la técnica de cuestionario o encuesta para la recolección de los datos, ellos mismos deben hacer la construcción del cuestionario. El profesor puede guiarlos en el tipo de preguntas que deberían realizar, para que no sea difícil de organizar los datos ya recolectados. Se adjunta un ejemplo de cuestionario para este tipo de actividad.

### ***Tú programa de televisión preferido***

**Introducción:** Los estudiantes de la sección \_\_\_\_\_ estamos realizando un trabajo de Estadística en la clase de Matemática, para lo cual le pedimos su colaboración. La información que nos facilite no será revelada en casos individuales.

A. Información general:

1. Sexo: ( ) Mujer ( ) Hombre

B. Información relacionada con los programas de televisión

2. ¿Cuántas horas al día miras la televisión?

a) Menos de 2

b) Más de 2 pero menos de 5

c) Más de 5

3. Qué tipo de programas te gustan más (marca UNA ÚNICA OPCIÓN):

<p>a) Deportes  b) Telenovelas  c) Teleseries  d) Películas  e) Documentales científicos e históricos  f) Otros: _____</p> <p>4. ¿Cuál es el nombre de tu programa de televisión preferido?  _____</p> <p style="text-align: center;"><i>Muchas gracias por su colaboración</i></p>
---

Seguidamente deben iniciar el proceso de organización de los datos. El profesor puede orientar a sus estudiantes para que realicen un cuadro donde se resuma la información obtenida por medio de los cuestionarios, como el siguiente:

Número de cuestionario	Grado	Edad	Cuántas horas mira TV al día	Tipo de programa	Programa preferido

Luego puede hacer preguntas generales (a todos los grupos) para guiarlos en la construcción de los cuadros estadísticos, tales como:

¿Cómo pueden mostrarle a los demás compañeros los resultados que obtuvieron con su cuestionario de manera resumida y clara?

Tomando en cuenta las variables contempladas en el cuestionario clasifíquenlas en cualitativas y cuantitativas ¿Cuáles se pueden representar en un cuadro? ¿Cuáles son las características de cada uno de esos cuadros? ¿Cuáles de las variables pueden relacionarse en un cuadro de doble entrada?, entre otras.

Se debe brindar el tiempo adecuado para que los estudiantes puedan recolectar y organizar los datos. En esta etapa el rol del docente es completamente activo, debe involucrarse con los jóvenes para orientar sus trabajos y percepciones; pero debe permitir la discusión entre ellos en relación con la búsqueda de soluciones.

### Discusión interactiva y comunicativa

En este momento, cada subgrupo de estudiantes expone ante el docente y sus compañeros los resultados que obtuvieron al administrar los cuestionarios.

La idea con la exposición de los resultados es que los estudiantes defiendan su trabajo ante los demás por lo que las representaciones tabulares deben contar con todos los elementos necesarios; a saber: título, fuente, nombre de cada uno de los ejes, escala y todo esto contextualizado a los datos obtenidos. Además, debe hacerse uso de frecuencias tanto relativas como absolutas.

En la exposición el profesor puede realizar preguntas relacionadas principalmente con la interpretación de los datos y representaciones utilizadas, por ejemplo:

- ¿Cuál es el programa preferido de las mujeres?
- ¿Cuál es el programa preferido de los hombres?
- ¿Qué tipo de programas son los más gustados por los compañeros?
- ¿Cuál es el programa que la menor cantidad de estudiantes entrevistados prefiere?

De manera similar para las otras características.

## Clausura o cierre

Al finalizar las exposiciones el docente puede hacer una recopilación de opiniones de los estudiantes para conformar los conceptos de recolección de información por medio de un cuestionario, características de un buen cuestionario; además, explicar mostrar a las y los estudiantes los procedimientos a seguir en la construcción de una representación tabular y los elementos necesarios en estas representaciones.

Como actividad de autoevaluación el docente podría asignarle a sus estudiantes que conformen un pequeño cuestionario y lo apliquen a sus vecinos para conocer su opinión en aspectos como: medidas de reciclaje, medidas para el ahorro del agua y la electricidad, etc. y que los estudiantes le presenten un reporte con los resultados obtenidos, donde el uso de cuadros, gráficas, así como la interpretación de la información esté presente.

## Reflexión

A esta recomendación metodológica se le podrían realizar algunas variantes y utilizarse en octavo grado. Por ejemplo, se puede dar mayor énfasis a los datos cuantitativos continuos en los cuestionarios sin predefinir las clases, para así realizar las distribuciones de frecuencias respectivas.

## Bibliografía

- Batanero, C. (2002). Los retos de la cultura estadística. Jornadas Interamericanas de Enseñanza de la Estadística, Buenos Aires, Argentina. Conferencia inaugural.
- Batanero, C. y Godino, J. (2002). Estocástica y su didáctica para maestros. En Godino, J. (2002). Matemática y su Didáctica para Maestros: Manual para el estudiante. Pág.753-761 <http://www.ugr.es/local/jgodino/edumat-maestros/>
- Batanero, C. (2001). Didáctica de la estadística. Granada, España: Grupo de Educación Estadística de la Universidad de Granada.
- Gómez, M. (1999). Elementos de Estadística Descriptiva. San José, Costa Rica: EUNED.
- Quirós, J. (2010). Contribución a una propuesta metodológica de la enseñanza de la Estadística en Octavo año de la Enseñanza General Básica. Tesis de Licenciatura . Escuela de Matemática, UCR.
- Fernández, C. y Fuentes, F. (1995). Curso de estadística descriptiva: teoría y práctica. España: Ariel.
- Meter, J., y Wasserman, W. (1973). Fundamentos de Estadística. Traducción de la tercera edición. Compañía Editorial Continental, S.A. (C.E.C.S.A).
- Quintana, C. (1996). Elementos de Inferencia Estadística. San José, Costa Rica: EUCR.

## Créditos

Esta unidad didáctica es parte del *Curso bimodal para el Tercer Ciclo: Enfoque de Resolución de problemas*, que forma parte del proyecto *Reforma de la Educación Matemática en Costa Rica*.

Este proyecto del Ministerio de Educación Pública de Costa Rica es apoyado financieramente por la Fundación Costa Rica-Estados Unidos de América para la Cooperación, y es ejecutado administrativamente por la Fundación Omar Dengo.

### **Autores**

Marianela Alpízar Vargas  
Edwin Chaves Esquivel  
Ana Lucía Alfaro Arce

### **Revisor**

Jonathan Espinoza González

### **Editor gráfico**

Miguel González Ortega

### **Director general del proyecto *Reforma de la Educación Matemática en Costa Rica***

Ángel Ruiz

### **Para referenciar este documento**

Ministerio de Educación Pública (2011). *Curso bimodal para el Tercer Ciclo: Enfoque de Resolución de problemas. Unidad didáctica Estadística*. San José, Costa Rica: autor.



Curso bimodal para el Tercer Ciclo: Enfoque de Resolución de problemas. Unidad didáctica Estadística por Ministerio de Educación Pública de Costa Rica se encuentra bajo una Licencia [Creative Commons Atribución-NoComercial-CompartirIgual 3.0 Unported](https://creativecommons.org/licenses/by-nc-sa/3.0/).