ELEMENTOS DE ESTADÍSTICA Y SU DIDÁCTICA A NIVEL BACHILLERATO

ERNESTO SÁNCHEZ SÁNCHEZ

SUBSECRETARÍA DE EDUCACIÓN MEDIA SUPERIOR

ELEMENTOS DE ESTADÍSTICA Y SU DIDÁCTICA A NIVEL BACHILLERATO

Ernesto Sánchez Sánchez

Investigador de

DME-Cinvestav

Ricardo Cantoral Uriza
Coordinador de la Serie

Primera edición, 2013
© Secretaría de Educación Pública, 2013
Subsecretaría de Educación Media Superior
Argentina # 28 Col. Centro Histórico, Del. Cuauhtémoc
México, Distrito Federal

ISBN: 978-607-9362-00-3

Impreso en México

Se permite la reproducción del material publicado previa autorización del editor. Los textos son responsabilidad de los autores y no reflejan, necesariamente, la opinión de la Subsecretaría de Educación Media Superior.

CONTENIDO

PRÓLOGO

Estimada profesora, estimado profesor:

Como parte de una estrategia de largo plazo para la profesionalización docente en el campo de las matemáticas, el Departamento de Matemática Educativa del Centro de Investigación y de Estudios Avanzados (Cinvestav) y la Subsecretaría de Educación Media Superior de la SEP, diseñaron un plan para elaborar estos materiales dirigidos a las y los profesores de Matemáticas del país. En un segundo momento, con la colaboración de la red de egresados de Matemática Educativa y el apoyo de la Sociedad Matemática Mexicana, llevaremos a cabo mesas, foros, seminarios, cursos y diplomados mediante un Plan Nacional para la Profesionalización Docente en las Matemáticas Escolares.

Quienes estamos interesados en el aprendizaje de las matemáticas no podemos reducir los conceptos a sus definiciones, ni limitar las experiencias didácticas a la repetición memorística de algoritmos y resultados. Aprender matemáticas no puede limitarse a la mera copia del exterior a través de resultados previamente elaborados, o digamos que, a su duplicado; sino más bien, es el resultado de construcciones sucesivas cuyo objetivo es garantizar el éxito ante una cierta situación de aprendizaje.

Una consecuencia educativa de este principio consiste en reconocer que tenemos todavía mucho que aprender al analizar los propios procesos de aprendizaje de nuestros alumnos; nos debe importar, por ejemplo, saber cómo los jóvenes del bachillerato operan con los números, cómo entienden la pendiente de una recta, cómo construyen y comparten significados relativos a la noción de función o proporcionalidad, o cómo se explican a sí mismos nociones de azar. Esta visión rompe con el esquema clásico de enseñanza según el cual, el maestro enseña y el alumno aprende. Estos textos se diseñaron para ayudar al docente a explorar y usarlos para una enseñanza renovada aprovechando las formas naturales en que los estudiantes razonan sobre matemáticas y sobre lo que aporta a este respecto la investigación en Matemática Educativa.

El papel del profesor en esta perspectiva es mucho más activo y propositivo, pues sobre él o ella recae más la responsabilidad del diseño y coordinación de las situaciones de aprendizaje. Actualmente se considera al profesor como un profesional reflexivo, que decide, diseña, aplica y experimenta estrategias de acción para lograr el aprendizaje de sus alumnos. De manera que aprender matemáticas no se reduce a recordar fórmulas, teoremas o definiciones para resolver problemas mediante la imitación de las explicaciones del profesor en clase o con apego a los métodos ilustrados en los textos escolares.

Los resultados de las pruebas nacionales de corte masivo, utilizadas con fines de investigación, permitirían saber cuáles conceptos y procesos requieren todavía adaptaciones progresivas con el fin de mejorar su aprendizaje. Si bien los últimos resultados de las pruebas de logro académico estandarizadas muestran un incremento en el porcentaje de la población estudiantil con resultados satisfactorios y un decremento en el complemento, aún falta mejorar la atención en algunas temáticas particulares.

Gracias a la labor que lleva a cabo el Departamento de Matemática Educativa del Cinvestav, a través de sus profesores, egresados e investigadores en formación, sabemos cuáles asuntos, de naturaleza transversal, resultan fundamentales para el aprendizaje de las matemáticas y de las ciencias, como puede ser el desarrollo del pensamiento y lenguaje variacional, la constitución de un lenguaje gráfico para las funciones, el desarrollo del pensamiento trigonométrico, el pensamiento proporcional y el pensamiento estadístico. Estos asuntos siguen siendo un reto de la mayor importancia para mejorar los aprendizajes entre los estudiantes del bachillerato mexicano.

Por esta razón, los cinco volúmenes de esta colección fueron pensados para el docente de matemáticas. Su lectura, análisis y discusión permitirá mejorar los procesos de aprendizaje matemático. Los títulos de los textos de la serie son los siguientes:

Vol. 1 - *Lenguaje gráfico de funciones. Elementos de precálculo*

- Rosa María Farfán Márquez

Vol. 2 - *Desarrollo del pensamiento trigonométrico*

- Gisela Montiel Espinosa

Vol. 3 - *Desarrollo del pensamiento y lenguaje variacional*
- Ricardo Cantoral Uriza

Vol. 4 - *La transversalidad de la proporcionalidad*

- Daniela Reyes Gasperini

Vol. 5 - *Elementos de estadística y su didáctica a nivel bachillerato*

- Ernesto Sánchez Sánchez

Según la profesora Régine Douady, *saber matemáticas* precisa de dos aspectos. Por un lado, se refiere a la disponibilidad funcional de nociones y teoremas matemáticos para enfrentar problemas e interpretar nuevas situaciones. En este proceso, dichas nociones y teoremas tienen un estatus de herramienta, en tanto que sirven para que alguien actúe sobre un problema en determinado contexto. Por otra parte, también significa identificar a las nociones y a los teoremas como parte de un cuerpo de conocimientos reconocidos socialmente. Es ahí donde se formulan definiciones, se establecen relaciones entre nociones mediante teoremas y se prueban las conjeturas adquiriendo entonces el estatus de objeto. Al adquirir ese estatus, están descontextualizados y despersonalizados para permitir su aprendizaje. Este proceso de descontextualización y despersonalización participa del proceso de apropiación del conocimiento. Por su parte, para un profesor enseñar se refiere a la creación de las condiciones que producirán la apropiación del conocimiento por parte de los estudiantes. Para éstos, aprender significa involucrarse en una actividad intelectual cuya consecuencia final sea la disponibilidad de un conocimiento con su doble estatus de herramienta y de objeto. Para que haya aprendizaje y enseñanza es necesario que el conocimiento sea un objeto importante, casi esencial, de la interacción entre el profesor y sus alumnos.

Ésta es pues la primera de una serie de iniciativas coordinadas para la mejora de la educación en el campo de las matemáticas del bachillerato. No me resta más que animarles a estudiar y discutir los materiales que ahora tienen en sus manos, el camino es largo, pero iremos juntos ...

Dr. Ricardo Cantoral Uriza

Jefe del Departamento

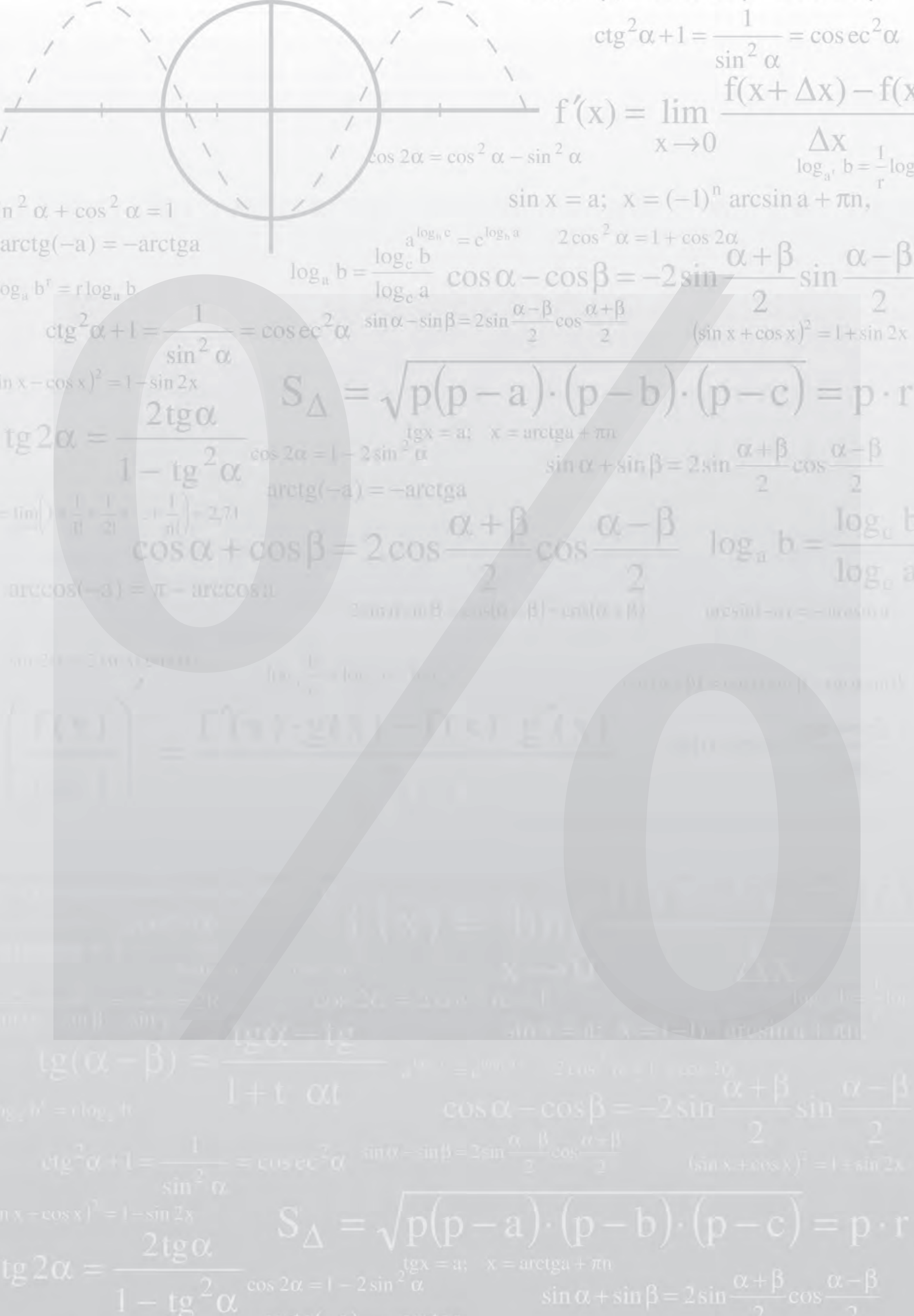
Matemática Educativa – Cinvestav

INTRODUCCIÓN

La estadística juega un papel fundamental en las sociedades actuales en las que se producen y utilizan grandes cantidades de información, pues tiene que ver con las formas en que se recogen, organizan y comunican conjuntos de datos y con la manera en que se analizan para hacer inferencias y predicciones, y para tomar decisiones. La estadística extiende su radio de influencia a todas las disciplinas científicas y sociales, ya que, como afirma Moore (2000), proporciona un método general útil para tratar datos, estimar su variación y el riesgo en situaciones de azar, de manera que es importante en la cultura de la mayoría de las personas. La caracterización de Moore no excluye la importancia de la estadística en la vida cotidiana, pues los individuos a lo largo de su vida requieren entender diversidad de información estadística que puede afectar sus vidas. Asimismo, los ciudadanos tienen que contender con fenómenos variables y azarosos frente a los cuales es necesario tomar decisiones racionales. Probablemente es el reconocimiento de todas estas razones lo que ha hecho que se incluya la estadística en el currículo de la educación preuniversitaria desde hace varias décadas.

Por otra parte, en los últimos años también ha surgido con fuerza la idea de que la educación debe asegurar la formación de ciudadanos competentes, en el sentido de que puedan hacer funcionar o utilizar de manera efectiva el conocimiento adquirido en la escuela en su vida personal, social y profesional. Sin embargo, hay evidencias de que esto es difícil de alcanzar, y de que los anteriores currículos, distintos al enfoque actual basado en el desarrollo de competencias, no propiciaban este tipo de educación. En gran medida, las evidencias mencionadas se derivan de los resultados obtenidos en las pruebas de evaluación nacionales e internacionales. En particular, cuando estos resultados han sido negativos, la sociedad los ha interpretado como deficiencias o bajos logros en los objetivos educativos nacionales. Con la intención de revertir esta situación, diversos autores han sugerido replantear los contenidos curriculares, reforzar la formación y actualización de profesores y promover el cambio de las prácticas docentes en las aulas.

En este trabajo se intenta proporcionar al profesor un instrumento que lo apoye para realizar una transición que lo conduzca hacia un enfoque de la enseñanza de la estadística basado en el desarrollo de competencias. Para hacerlo, hemos elegido tres temas tradicionales del currículo: gráficas, medidas de tendencia central y medidas de variación. Proponemos en el primer capítulo un acercamiento clásico del tema. En el segundo, se exponen algunos de los resultados más relevantes que se han propuesto desde la didáctica de la estadística con relación a los temas elegidos. En este capítulo queda en evidencia que el enfoque clásico no es suficiente para generar los conocimientos, habilidades y actitudes que se requieren para que los estudiantes sean usuarios competentes de la estadística. En el tercer capítulo se ofrecen algunos de los aspectos importantes que permiten entender hacia dónde conviene conducir la enseñanza de la estadística. Específicamente, se revisan cuáles son las relaciones entre la educación estadística y la propuesta de un currículo por competencias. Se argumenta que la estadística, por su propia naturaleza, es especialmente adaptable a dicha propuesta y que la forma de concretarla en el aula es mediante la enseñanza basada en proyectos. El capítulo 4 consiste en un ejemplo de un proyecto propuesto por Batanero, Díaz y Gea (2012), en el que se puede observar la potencialidad de este método de enseñanza.



1. GRÁFICAS, CENTROS Y DISPERSIÓN

LAS ESCALAS DE MEDIDA

La escala nominal. Los datos están sobre una escala nominal si son nombres o etiquetas con los que se indica que tienen un atributo. Por ejemplo, en un censo la variable "sexo" es nominal, no importando que los sexos se indiquen con números, por ejemplo "0" para "hombre" y "1" para "mujer"; lo mismo los valores 0 y 1 son sólo etiquetas. Ni el orden ni las operaciones que se pueden hacer con éstos, o cualesquiera otros números, tienen significado alguno cuando son nominales.

La escala ordinal. Los datos están sobre una escala ordinal si son tales que pueden ser ordenados y dicho orden tiene significado. Por ejemplo, los meses del año están a una escala ordinal, pues están ordenados de acuerdo con su posición en el año, tiene sentido decir que enero está antes de febrero y éste antes que marzo, etc. En ciertas condiciones puede ser conveniente asociar los meses del año con los números del 1 al 12, pues el orden de éstos tiene significado. En una escala ordinal la diferencia entre dos elementos o su distancia no necesariamente tiene significado; por ejemplo, en la anterior asignación, la distancia de 1 a 6 es 5, esto no quiere decir que la distancia de "enero"(1) a "junio" (6) es "mayo" (5).

La escala de intervalos. Los datos están en una escala de intervalos si tanto el orden como la distancia entre elementos son significativos. El ejemplo más claro de datos que están sobre una escala de intervalos son los datos que provienen de mediciones de la temperatura en grados centígrados, los cuales se pueden ordenar y el intervalo entre dos valores es significativo, ya que representa la temperatura que se necesita aumentar a la primera temperatura para llegar a la segunda. En las escalas de intervalos el cero no es significativo, es convencional. La temperatura de cero grados centígrados no significa ausencia de calor.

La escala de razón. Los datos están sobre una escala de razón si pueden ordenarse, los intervalos (o diferencias) son significativos y el cero también es significativo. Estas propiedades de los datos de una escala de razón hacen significativas las razones entre los datos. Las medidas de distancia están en una escala de razón, ya que se pueden ordenar, la diferencia entre dos distancias es significativa y una distancia cero representa cero desplazamientos; si la razón de dos distancias es ½, significa que una distancia es la mitad de la otra.

TABLAS DE DATOS Y GRÁFICAS

La primera organización que se le suele dar a los datos, una vez que son recogidos, consiste en indicar el conjunto de medidas observadas de la variable en estudio junto con el número de veces en que cada medida u observación ocurre. Los arreglos que ofrecen dicha información se llaman *distribuciones de frecuencia*. En la figura 1 se muestra una distribución en forma de tabla y en forma de diagrama de columnas sobre mortalidad por cáncer.

Panorama Internacional de la Mortalidad por cáncer

**Mortalidad por cáncer.
Panorama Internacional, 2007**

Tipos de cáncer	2007
Pulmón	1,400,000
Mama	549,000
Colon	677,000
Estómago	866,000
Hígado	653,000

Fuente: Informe de Cáncer Organización Mundial de Salud.

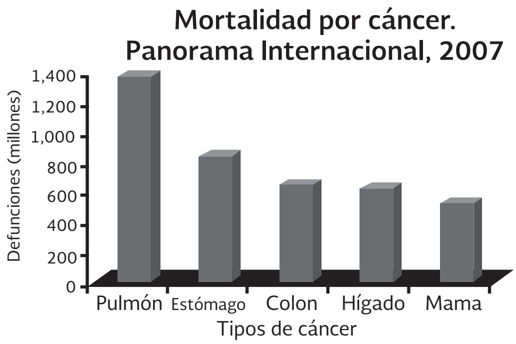


Figura 1. Tabla de frecuencia y diagrama de columnas de los tipos de cáncer más frecuentes por Mortalidad.
Fuente: http://www.dgepi.salud.gob.mx/2010/PDFS/PUBLICACIONES/MONOGRAFIAS/PEPID_ TUMORES_MALIGNOS_MEX_2010.pdf (consultada el 07/07/13).

Gráficas básicas de estadística

Si se quiere presentar en una gráfica los datos de una tabla con una sola fila (o columna) de datos, sólo se puede elegir entre los siguientes cinco tipos básicos:

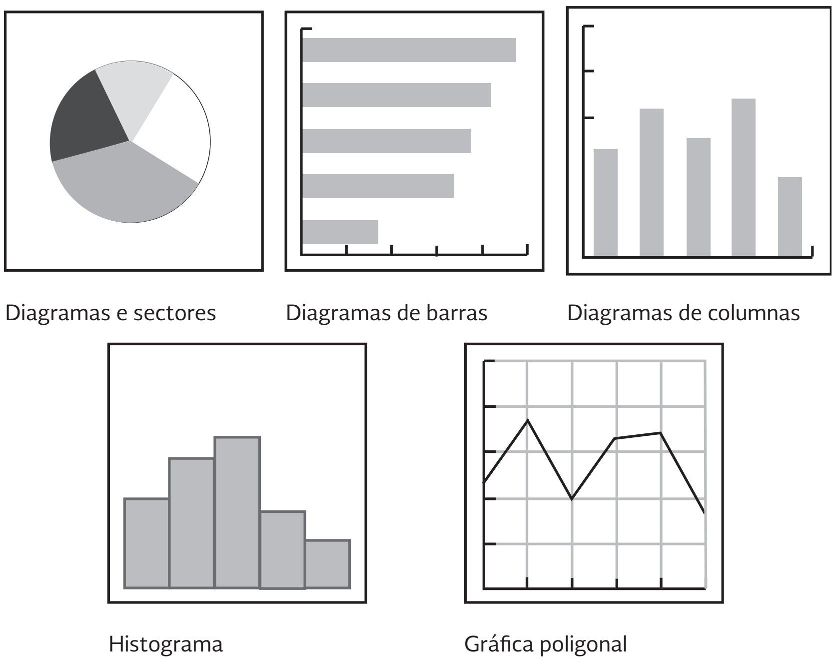


Figura 2. Tipos de gráficas.

Hay dos criterios generales para elegir el tipo de gráfica para presentar los datos:

Diagrama de sectores

Este diagrama resulta conveniente cuando la variable es nominal u ordinal, con no más de seis valores distintos; es importante tener en cuenta que los datos deben cubrir una totalidad (100%). Cada sector es proporcional a la frecuencia del valor de la variable que representa. En el ejemplo de la figura 3, se presenta un diagrama de sectores del número de municipios gobernados por cada partido (en 2011):

Partido	Municipios
PRI	15210
PAN	473
PRD	322
Otros	135

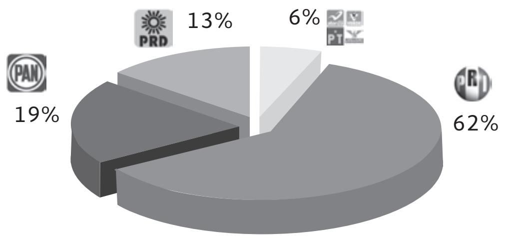


Figura 3. Tabla de frecuencia de municipios gobernados por los partidos.

Fuente: http://www.fenamm.org.mx/site/index.php?option=com_conten&id=187&Itemid=123(Consultada en:07/07/213).

Diagrama de barras

Este tipo de diagrama también es una opción conveniente cuando la variable es nominal y, a diferencia del diagrama de sectores, no pierde claridad aun cuando hay un número grande de valores de la variable. Por ejemplo,

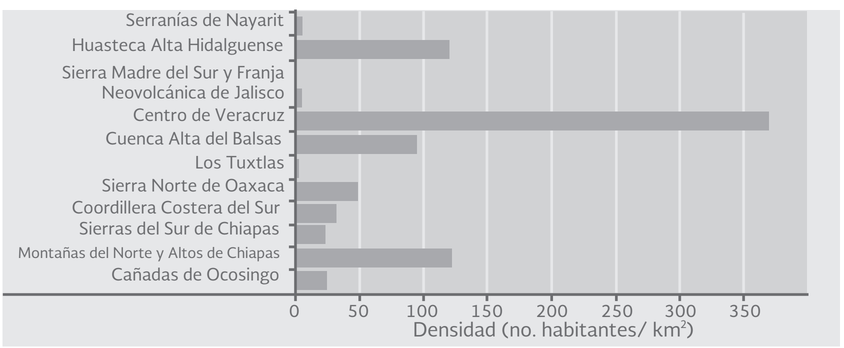


Figura 4. Densidad poblacional (Núm. de habitantes por kilómetro cuadrado) en los bosques de montaña de México y su área de influencia (3 Km).
Fuente: Conabio (2010). El bosque mesófilo de montaña en México: Amenazas y oportunidades para su conservación y manejo sostenible. México: Gobierno Federal.

Diagrama de columnas

Se prefiere un diagrama de columnas cuando la variable independiente tiene valores sobre una escala ordinal. A diferencia del diagrama de barras, el de columnas no admite una reordenación de los valores de la variable independiente, ya que éstos tienen un orden natural. Por ejemplo, supongamos que la variable independiente toma los valores de los meses del año, resultaría inapropiado seguir un orden diferente al de la dirección del tiempo.

Veamos la gráfica “Evolución del número de bibliotecas”, que se presenta a continuación. Observemos que los años que van de 1980 a 1997 constituyen la variable independiente, la cual ahora se representa en el eje horizontal; la variable dependiente “número de bibliotecas” se representa en el eje vertical. Este diagrama tiene algunos elementos decorativos, como presentar las barras con efecto de tercera dimensión y el jaspeado de la superficie de fondo detrás de las barras. Por lo demás, las líneas horizontales y las divisiones en los ejes son elementos necesarios de una buena gráfica de columnas; los diagramas en los que se omite alguno de éstos pueden ser aceptables, pero baja su calidad.

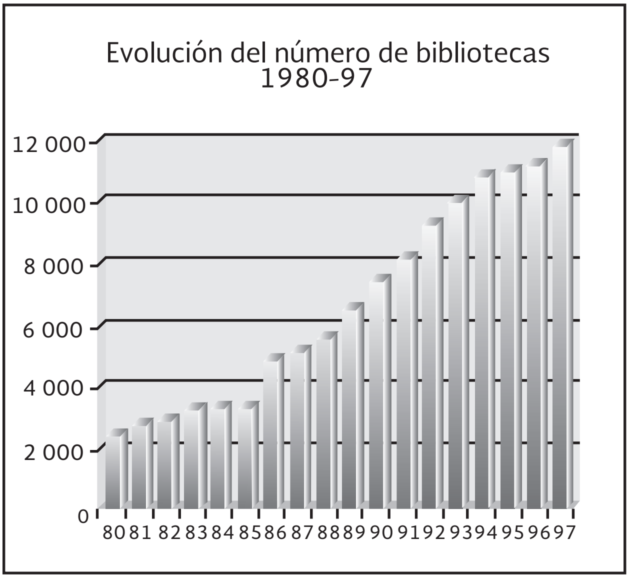


Figura 5: Evolución del número de bibliotecas de 1980-1997.
Fuente: INEGI, Anuario Estadístico, Estados Unidos Mexicanos, 1999.

Nota: En muchos textos se refieren a los diagramas de columnas también como diagramas de barras. Lo importante no es el nombre, sino la elección de la forma adecuada para mejorar la comunicación.

Histograma y gráfica poligonal

Estas formas de presentación de la información se prefieren cuando la variable tiene valores sobre una escala de intervalo o de razón. Esto significa que el eje horizontal es una recta numérica en la que cada punto corresponde a un número significativo. El histograma se compone de una serie de rectángulos, cada uno de los cuales tiene como base un intervalo de clase y, generalmente, la altura de cada rectángulo representa la frecuencia con la que se presentan datos contenidos en el intervalo que se representa.

El histograma no es simplemente un diagrama de columnas con las barras juntas, ya que en este tipo de gráfica tiene sentido el ancho de los rectángulos e incluso en algunos diseños los intervalos puede no ser iguales. Mientras que en un diagrama de columnas el espacio entre las columnas y su ancho es convencional, en el histograma el ancho de los rectángulos porta información.

Ejemplo:

Tabla 1. Embarazadas según aumento de peso durante la gestación, Centro de Salud "Luis Pasteur",1991

Aumento de peso*	Número	Porcentaje
6 a 8	68	53.13
9 a 10	24	18.75
11 a 12	16	12.50
13 a 14	20	15.62
Total	128	100.00

* En kilogramos enteros.

Para hacer el histograma de la información que está representada en la tabla se debe observar que la variable independiente "Aumento de peso" es una variable continua, pero si se dibujan los intervalos tal como se miran en la tabla obtendríamos espacios vacíos:

Figura 6. Ejemplo de histograma.

Pero es fácil darse cuenta que no es poco probable que haya habido personas que aumentaron alrededor de 8.25 kg o 8.75 kg, es decir, que seguramente se obtuvieron valores que corresponden a los "hoyos" de la recta dibujada arriba. Se puede suponer que quienes colectaron los datos llevaron a cabo un redondeo. Así, aumentos de pesos cerca de 8.25 lo registraron como 8 y 8.75 como 9. Bajo este supuesto (que es el más natural en estos casos), los "hoyos" se "rellenan" dividiéndolos entre dos y adhiriéndolos a los intervalos adyacentes. Al hacer lo anterior se obtienen nuevos intervalos que llenan el segmento en el que varían los datos. Los puntos medios de cada intervalo se llaman marcas de clase. A continuación se presentan los intervalos y su marca de clase y más abajo el histograma correspondiente:

(5.5,8.5) su marca de clase es 7

(8.5,10.5) su marca de clase es 9.5

(10.5,12.5) su marca de clase es 11.5

(12.5,14.5) su marca de clase es 13.5

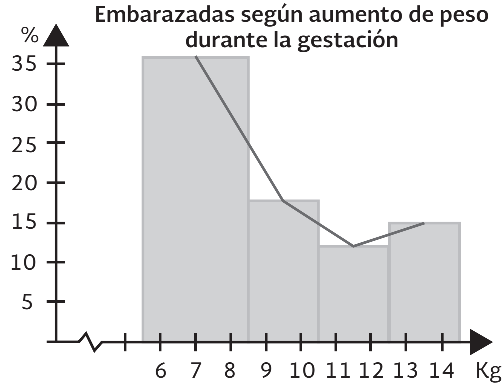


Figura 7. Ejemplo de histograma.

ESTADÍSTICOS IMPORTANTES

A cada conjunto de datos se le pueden asociar valores que dicen algo acerca de sus elementos, a esos valores se les llama *estadísticos*. Por ejemplo, si los datos de una colección están sobre una escala al menos ordinal, su valor mínimo es un estadístico; éste proporciona información de todo el conjunto, a saber, que cualquier valor de la colección es mayor o igual que él. Los estadísticos más conocidos y utilizados son la moda, la mediana y la media o promedio.

> La *moda* es el dato o puntuación que aparece en la lista con mayor frecuencia.

La moda es el estadístico más importante de una lista de datos cuando éstos están sobre una escala nominal. En el ejemplo 3.1 lo importante es saber qué canción ganó con un número mayor de llamadas. En el segundo ejemplo la moda de los datos es 49, pero este valor no es tan importante, en su lugar, es más importante el intervalo de moda, es decir, el intervalo que tiene mayor número de puntuaciones: (51–60); a ese intervalo pertenece el resultado que obtuvo Juan.

La mediana de una lista de datos es el número que divide la lista en dos partes, de tal manera que la mitad de los datos son menores o iguales a la mediana y la otra mitad son mayores o iguales a ella.

Por ejemplo:

24	29	31	31	35	37	38	39	42	43
45	45	47	48	49	49	49	49	50	51
51	52	53	53	53	54	54	55	56	56
57	58	59	63	63	64	64	66	68	69
69	70	73	73	74	75	78	81	84	90

La mediana es, entonces, el valor intermedio entre 53 y 54, es decir: 53.5. Observe que en este caso la mediana no coincide con ninguno de los valores de la lista. De paso podemos agregar con relación a la preocupación de Juan que la puntuación que obtuvo está por arriba de la mediana.

La *media* o *promedio* de un conjunto de datos numéricos es el número que se obtiene de sumar los datos y dividirlos entre el número total de ellos.

La media es el estadístico más utilizado cuando las puntuaciones están sobre una escala de intervalo o de razón. En la escuela es muy conocido y utilizado el promedio, que no es otra cosa que la media de las calificaciones obtenidas. En el lenguaje de algunas especialidades se habla de la estatura promedio, del peso medio, etc., esos valores corresponden con el concepto de media que acabamos de definir.

Como los valores del ejemplo anterior están sobre una escala de razón, la media para ellos tiene sentido, así que podemos obtenerla. Sumemos todos los valores y el resultado de la suma lo dividimos entre 50:

$$\text{Suma de todos los valores} = 2766$$

$$\text{Media} = \frac{2766}{50} = 55.32$$

Media de datos agrupados

En ocasiones no se cuenta con la lista de datos sino con una tabla de frecuencias. En esos casos también se puede calcular la media multiplicando cada marca de clase por su frecuencia, sumar todos los productos y finalmente dividirlo entre la suma de las frecuencias.

Para ejemplificar cómo se calcula la media de datos agrupados veamos la tabla que corresponde a los datos del ejemplo:

Intervalo	Marcas	Frec.
20 - 29	//	2
30 - 39	//////	6
40 - 49	//////////	10
50 - 59	///////////////	15
60 - 69	////////	8
70 - 79	//////	6
81 - 90	///	3
91 - 100		0
Total		50

Las marcas de clase son los puntos medios de cada intervalo, en el caso anterior son:

25 35 45 55 65 75 85 95

cada uno de estos valores se multiplica por su frecuencia y se suman:

$25\times2 + 35\times6 + 45\times10 + 55\times15 + 65\times8 + 75\times6 + 85\times3 + 95\times0 = 2760$

Entonces se divide este número entre el total de frecuencias que es 50, obteniéndose:

$$\text{Media}= \frac{2766}{50} = 55.32$$

Debemos observar que la media de los datos agrupados no es exactamente igual a la obtenida cuando los datos están sin agrupar, pero es muy aproximada; en el caso anterior la diferencia es sólo 0.11.

Mediana

Para ejemplificar el caso de la mediana en datos agrupados en frecuencias consideremos los datos de la tabla anterior (edad de las madres al tener su primer hijo). Señalamos anteriormente que cuando los datos se agrupan pierden su identidad y que la marca de clase asume la representatividad de los datos del intervalo. En este caso tendríamos 17 mamás que tuvieron su primer hijo a los 17 años, 15 que lo tuvieron a los 24 años y así sucesivamente. Como se tienen 51 datos ordenados, la mediana se encuentra en la posición 26 debido a que

$$P_m = \frac{n+1}{2} = \frac{51+1}{2} = 26$$

Es decir, la mediana se encuentra en el segundo intervalo dado que el acumulado al primer intervalo son 17 datos, mientras que el acumulado en el segundo intervalo son 32 datos. Lo anterior nos conduce a que la mediana se encuentra entre el valor 21 y 27, que son los límites del intervalo que contiene a la mediana.

Una expresión que nos permite calcular la mediana es la siguiente:

$$\text{Mediana} = L_m + \left[\frac{\frac{n}{2} - \sum f}{f_m}\right] c$$

Donde:

L_m: Límite inferior del intervalo que contiene a la mediana.

n: Total de datos.

c: Amplitud del intervalo que contiene a la mediana.

f_m: Frecuencia del intervalo que contiene a la mediana.

$\sum f$: Suma de frecuencias anteriores al intervalo que contiene a la mediana.

Como se puede ver, para utilizar la fórmula es muy importante que primeramente se identifique el intervalo que contiene a la mediana, ya que todos los elementos hacen referencia a él.

Para el caso que estamos analizando se tiene lo siguiente:

Intervalo que contiene a la mediana: 21-27

$L_m = 21$

$n = 51$

$c = 7$

$f_m = 15$

$\sum f = 17$

Sustituyendo:

$$\text{Mediana} = L_m + \left[\frac{\frac{n}{2} - \sum f}{f_m}\right] c = 21 + \left[\frac{\frac{51}{2} - 17}{15}\right] 7$$

$$\text{Mediana} = 24.96 \text{ años}$$

Obsérvese la similitud que hay en este caso entre media aritmética y la mediana, dado que los datos siguen una distribución aproximadamente simétrica, lo cual se puede ver en el siguiente histograma.

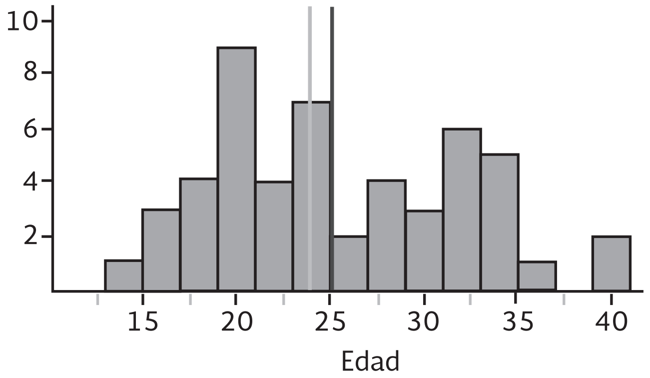


Figura 8. Histograma que muestra media y mediana.

MEDIDAS DE VARIABILIDAD O DISPERSIÓN

Las medidas de tendencia central no son suficientes para una adecuada descripción de los datos, ya que sólo nos indican el valor central alrededor del cual se encuentran los datos. Es necesario conocer otras medidas que nos indiquen el grado de dispersión o variabilidad de los datos para poder tener una idea más precisa de su comportamiento. Para comprender mejor lo anterior consideremos el caso de dos alumnos que obtuvieron las siguientes calificaciones en un semestre:

Alumno 1:	7	7	8	9	9
Alumno 2:	6	7	8	9	10

En ambos casos la media aritmética y la mediana son iguales a 8. Sin embargo, las calificaciones del segundo alumno son más variables que las del primero (ver figura 9). Así que a pesar de que ambos tuvieron el mismo promedio, en términos estadísticos ambos conjuntos de datos no tienen el mismo comportamiento.

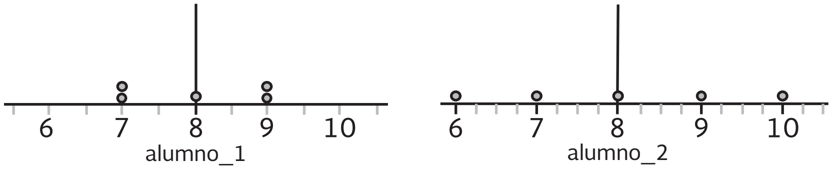


Figura 9. Dos conjuntos de datos con igual centro pero diferente variabilidad.

Entonces, en el análisis de datos es importante conocer el centro de un conjunto de datos, pero también se requiere conocer qué tan lejos del centro se encuentran los datos. En los siguientes párrafos abordaremos diferentes medidas que nos permitirán evaluar la variabilidad en los datos.

Rango. La medida de variabilidad más sencilla que existe es el rango. Representa la distancia entre el dato más pequeño y el más grande del conjunto. Esto es:

$$\text{Rango} = \text{dato mayor} - \text{dato menor}$$

Los datos que se muestran a continuación representan los niveles máximos de monóxido de carbono en la zona noroeste del Área Metropolitana de la Ciudad de México en la primera semana del mes de febrero de 2003 (Fuente: Centro de Control de la Red Automática de Monitoreo Atmosférico).

16 24 24 42 25 42 36

$\longleftarrow$ Rango $\longrightarrow$

$$\text{Rango} = 24 - 16 = 26$$

Obsérvese que el rango depende sólo de los datos extremos, por lo que resulta muy sensible a datos atípicos. Además no considera al centro de los datos, que debe ser la referencia respecto a la cual se mide la variabilidad. Debido a ello tiene poca utilidad. Ejemplos de sus aplicaciones pueden ser cuando se reportan temperaturas máximas y mínimas, índices máximos y mínimos de contaminación o en cartas de control de calidad de un producto, donde se reportan los extremos de la variable que se está midiendo durente un periodo. En todos los casos anteriores el rango proporciona una visión global de la variabilidad.

Desviación estándar y varianza

La desviación estándar es la medida de variabilidad más conocida y utilizada. Nos proporciona un valor típico que nos describe qué tan lejos se encuentran los datos de la media aritmética. Un elemento clave para entender la desviación estándar es el de desviación. Entenderemos por desviación la distancia de un dato respecto a la media. En un conjunto existen muchos datos, por lo tanto tendremos para cada uno de ellos una desviación.

La desviación de un dato respecto a la media se representa de la siguiente manera:

$$\text{desviación} = x - \bar{x}$$

Donde:

x: Es un dato del conjunto

$\bar{x}$: Es la media aritmética del conjunto de datos

De la expresión anterior se deriva que la desviación puede ser negativa o positiva, dependiendo si el dato es menor o mayor a la media aritmética. Para mostrar lo anterior consideremos una gráfica de puntos con los datos de la contaminación por ozono utilizados en el cálculo del rango:

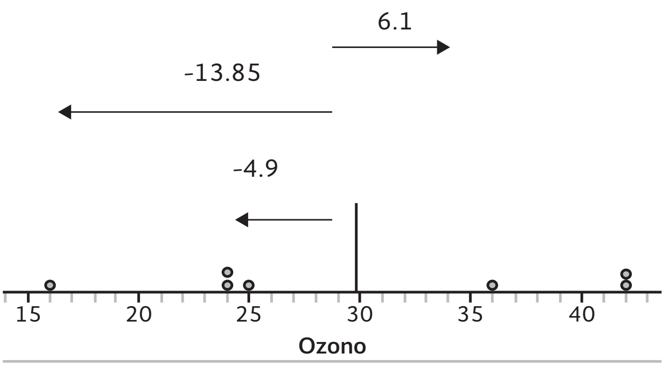


Figura 10. Diagrama de puntos con la media de los datos y algunas desviaciones.

En la sección anterior señalamos que la media aritmética constituía el punto de balance o equilibrio de un conjunto de datos. Ahora que introducimos el término de desviación, podemos ver que ello implica que la *suma de las desviaciones de los datos respecto a la media es igual a cero*. Ésta es sin duda una importante propiedad de la media. Simbólicamente se expresa de la siguiente manera:

$$\sum_{i=1}^{n}(x_1 - \bar{x}) = 0$$

Si calculamos las desviaciones para los puntajes de ozono, veremos que se cumple lo anterior.

Puntajes de ozono	Media	Desviaciones
16	29.85	16−29.85=-13.85
24	29.85	24−29.85=-5.85
24	29.85	24−29.85=-5.85
42	29.85	42−29.85=12.15
25	29.85	25−29.85=-4.85
42	29.85	42−29.85=12.15
36	29.85	36−29.85=6.15
		$\sum \text{Desviaciones} = 0$

En la búsqueda de una expresión para calcular la desviación estándar de un conjunto de datos, introduciremos el concepto de desviación cuadrática, el cual consiste en elevar al cuadrado las desviaciones. Ahora todas las desviaciones serán positivas. Esto es:

$$\sum_{i=1}^{n}(x_1 - \bar{x})^2 = (-13.85)^2 + (-5.85)^2 + (-5.85)^2 + (12.15)^2 + (-4.85)^2 + (12.15)^2 + (6.15)^2$$

$$\sum_{i=1}^{n}(x_1 - \bar{x})^2 = 616.85$$

El resultado anterior representa la suma de las desviaciones cuadráticas de todos los datos. Si dividimos dicho resultado entre el total de datos, estaremos calculando el promedio de las desviaciones cuadráticas. A este resultado se le denomina varianza, y es otra medida de variabilidad de los datos.

$$\frac{\sum_{i=1}^{n}(x_1-\bar{x})^2}{n}=\frac{616.85}{7}=88.1$$

Entonces, la varianza viene siendo el promedio (media aritmética) de las desviaciones cuadráticas de un conjunto de datos.

Si extraemos la raíz cuadrada al resultado anterior, se tendrá el promedio (media aritmética) de las desviaciones. A dicho resultado se le conoce como desviación estándar.

$$\sqrt{\frac{\sum_{i=1}^{n}(x_1-\bar{x})^2}{n}}=9.39$$

A partir de un conjunto de datos hemos explicado el proceso que se sigue para calcular la varianza y la desviación estándar. Ahora es necesario definir su simbología, para lo cual haremos dos distinciones importantes:

a) Cuando los datos provienen de una población:

La desviación estándar se representa mediante la letra σ y se expresa de la siguiente manera:

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}(x-\mu)^2}{N}}$$

Por su parte, la varianza, que es el cuadrado de la desviación estándar, se expresa:

$$\sigma^2=\frac{\sum_{i=1}^{n}(x-\mu)^2}{N}$$

b) Cuando los datos provienen de una muestra:

La desviación estándar se representa mediante la letra *S* y la varianza mediante S^2, respectivamente.

$$S=\sqrt{\frac{\sum_{i=1}^{n}(x-\bar{x})^2}{n-1}}$$

$$S^2=\frac{\sum_{i=1}^{n}(x-\bar{x})^2}{n-1}$$

Observaciones:

- μ es la media de la población.
- $\bar{x}$ es la media de la muestra.
- n es el total de datos en la muestra.
- N es el total de datos en la población.
- En el caso de la muestra el denominador es $n-1$ en el lugar de n.

La razón de dividir entre *n–1* y no entre *n* en el caso de la muestra, es porque se ha determinado que de esta manera se puede estimar con mayor precisión la media de una población en situaciones de inferencia estadística.

ACTIVIDADES Y PROBLEMAS

1. Observa la tabla y el diagrama que se presentan a la derecha y responde lo que se pide en cada inciso.

GOBIERNO Y POLÍTICA

DATOS GENERALES	
CONCEPTO	**2000**
Entidades federativas	32
Municipios	2,443
Senadores	128
Diputados federales	500
Esrados gobernados por el PRI	19
Esrados gobernados por el PAN	7
Esrados gobernados por el PRD	4
Esrados gobernados por el coalición	2
Ciudadanos en el padrón electoral (millones)	59.58

POBLACIÓN GOBERNADA POR PARTIDO (ÁMBITO MUNICIPAL), 2000 (%)

PAN	PRI	PRD	OTROS
35.81	44.42	17.32	2.45

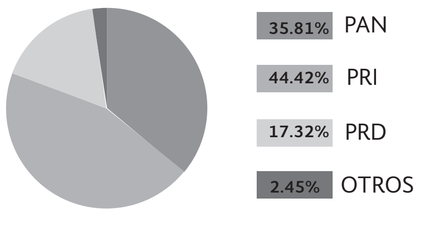


Figura 11. Representación de los datos de Gobierno y política.
Fuente: Aguayo, S., 2000, *El Almanaque Mexicano*, Grijalbo, México.

a) ¿Cuál es el título general del cuadro?

b) ¿Cuál es el título del diagrama?

c) ¿Qué tipo de variable se mide?

d) ¿Qué representan los números dados en porcentajes en el diagrama de sectores?

e) Redacta dos enunciados en los que expreses una opinión sugerida por los datos del diagrama.

f) ¿Cuál es la fuente?

2. Observa la siguiente tabla. Diseña un diagrama de sectores. Determina cuál es la variable independiente y cuál la dependiente

Parque vehicular, 1998

Automóviles	67.7%
Camiones de carga	31.5%
Autobuses	0.8%

Fuente: Aguayo, 2000, El Almanaque Mexicano, Grijalbo, Mexico.

3. Con los datos que se presentan en la tabla "Distribución de casos de cáncer en hombres y mujeres" diseña un diagrama de barras para cada uno de los sexos. Antes responde las preguntas.

a) ¿Cuál es la variable independiente para cada diagrama?

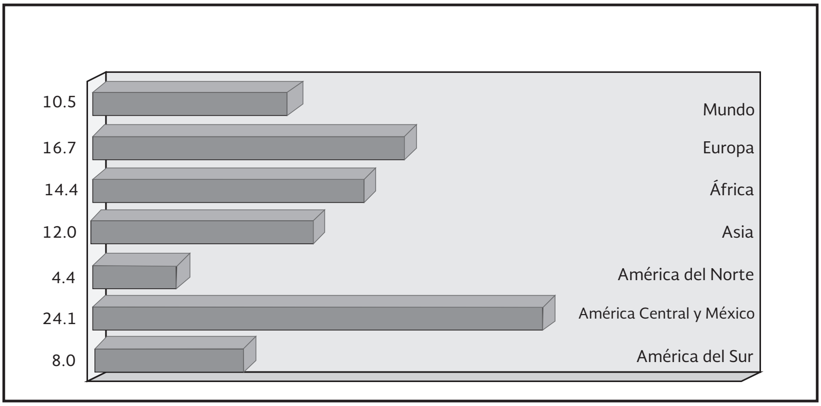
b) ¿Cuál es la variable dependiente para cada diagrama?

c) ¿Por qué no es conveniente representar la información en diagramas de sectores?

DISTRIBUCIÓN DE CASOS DE CÁNCER EN HOMBRE Y MUJERES,1995			
MUJERES			
LUGAR	**LOCALIZACIÓN**	**CASOS**	**%**
1	Cuello del útero	15,749	33.2
2	Mama	7,791	16
3	Ovario	1,684	3.5
4	Cuerpo del útero	1,432	3.0
5	Estómago	1,258	2.7
6	Glándula tiroides	1,211	2.6
7	Glándula linfáticos	1,179	2.5
8	Tejidos blandos	1,080	2.3
9	Vesícula biliar y vías intrahepáticas	763	1.6
10	Colon	728	1.5
HOMBRES			
LUGAR	**LOCALIZACIEÓN**	**CASOS**	**%**
1	Próstata	3,674	14.2
2	Estómago	1,620	6.3
3	Ganglios linfáticos	1,566	6.1
4	Tejidos blandos	1,332	5.2
5	Testículos	1,233	4.8
6	Tráquea, bronqueos y pulmón	1,139	4.7
7	Vejiga urinaria	1,136	4.7
8	Laringe	842	3.3
9	Encéfalo	702	2.7
10	Riñón y otros órganos urinarios	660	2.6

Fuente: Registro Hispatológico de Neoplasias en México, 1999.

4. Observa el siguiente diagrama de barras. Indica dos aspectos en los que el diagrama puede ser mejorado, teniendo en cuenta las recomendaciones para elegir el tipo de gráfica o diagrama que se dieron en el texto.



Tierras deterioradas a nivel mundial y por regiones, 1945 - 1990.
(Porcentaje de tierras fértiles que sufrieron degradación grave o moderada).
Fuente: Emily T. Smith y otros, *Growth vs. Environment. The push for sustainable development*, International Business Week/International Soil Reference & Information Centre, mayo 1992.

5. Los datos de la siguiente tabla corresponden a la presión sanguínea sistólica de 278 hombres adultos saludables.

a) Determina los intervalos de clase y redefine los intervalos de manera que se cubra la recta.
b) Diseña el histograma correspondiente.

Presión sanguínea	Frecuencia
100-104	2
105-109	6
110-114	13
115-119	25
120-124	31
125-129	47
130-134	53
135-139	42
140-144	30
145-149	17
150-154	8
155-159	4

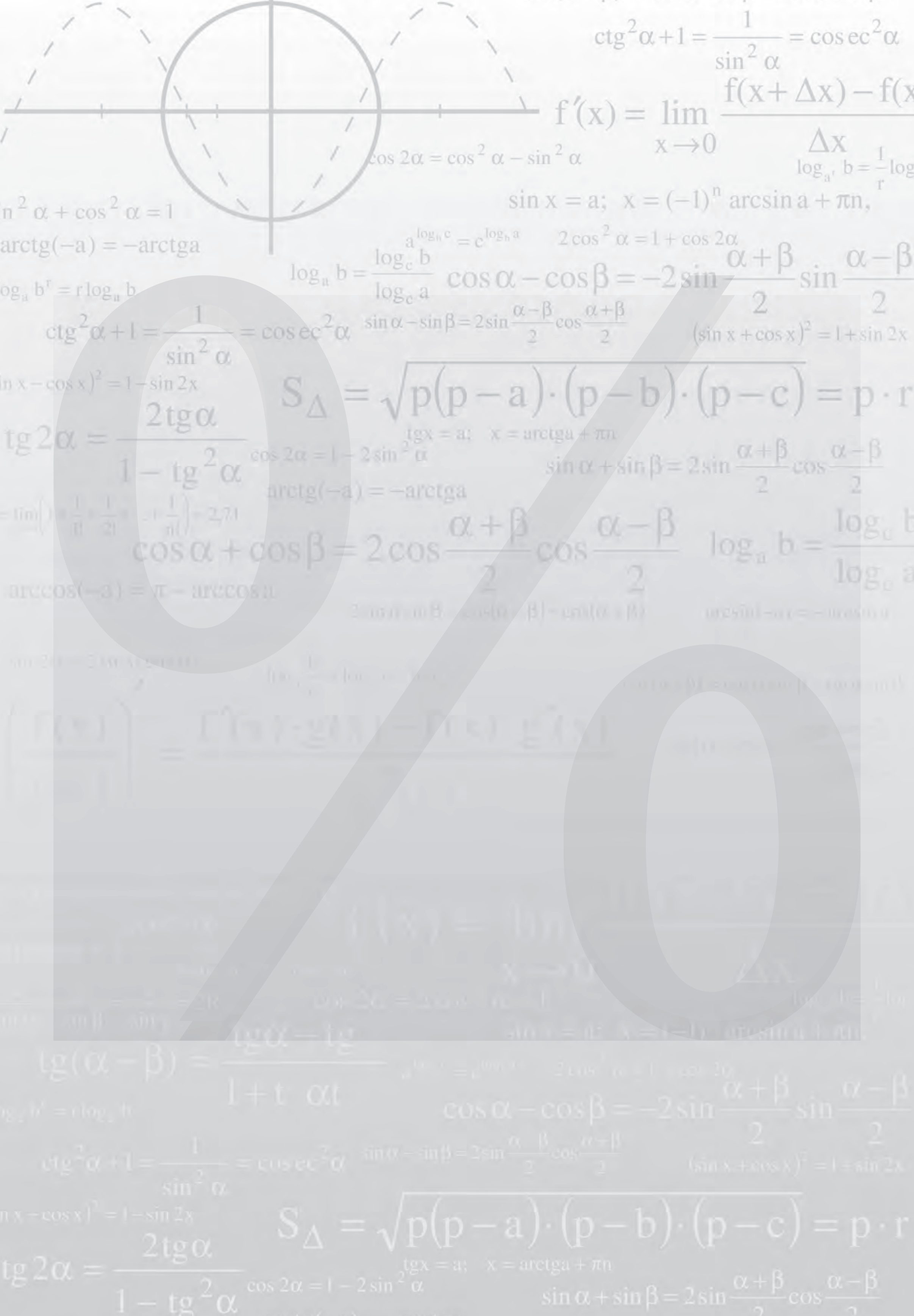
c) Traza su polígono de frecuencia.

6. Considera las calificaciones de los siguientes dos alumnos:

Alumno 1:	7	7	8	9	9	$x = 8$
Alumno 2:	6	7	8	9	10	$x = 8$

a) Por simple inspección: ¿cuál de los dos conjuntos de datos tiene mayor variabilidad?

b) Calculen la desviación estándar y verifique que coincide con su respuesta al inciso anterior.



$\text{ctg}^2\alpha + 1 = \frac{1}{\sin^2\alpha} = \text{cosec}^2\alpha$
$f'(x) = \lim_{x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$
$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$
$\sin x = a; \quad x = (-1)^n \arcsin a + \pi n,$
$\sin^2\alpha + \cos^2\alpha = 1$
$\text{arctg}(-a) = -\text{arctga}$
$a^{\log_b c} = c^{\log_b a}$
$2\cos^2\alpha = 1 + \cos 2\alpha$
$\log_a b = \frac{\log_c b}{\log_c a}$
$\cos\alpha - \cos\beta = -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$
$\log_a b^r = r\log_a b$
$\sin\alpha - \sin\beta = 2\sin\frac{\alpha-\beta}{2}\cos\frac{\alpha+\beta}{2}$
$(\sin x + \cos x)^2 = 1 + \sin 2x$
$(\sin x - \cos x)^2 = 1 - \sin 2x$
$S_\Delta = \sqrt{p(p-a)\cdot(p-b)\cdot(p-c)} = p \cdot r$
$\text{tg}\,2\alpha = \frac{2\text{tg}\alpha}{1 - \text{tg}^2\alpha}$
$\text{tg}x = a; \quad x = \text{arctga} + \pi n$
$\cos 2\alpha = 1 - 2\sin^2\alpha$
$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$
$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$
$\arccos(-a) = \pi - \arccos a$

2. ESTUDIOS EDUCATIVOS SOBRE GRÁFICAS, CENTROS Y DISPERSIÓN

El análisis de datos se apoya fuertemente en las representaciones gráficas debido a que los datos cuantitativos desplegados en forma visual ponen en evidencia algunas de sus características relevantes, que no pueden ser percibidas si sólo se observa la lista de datos o si se revisan los datos uno a uno. Esta potencia de las representaciones gráficas es muy útil para la comunicación en nuestra sociedad altamente tecnificada. Sin embargo, la lectura e interpretación adecuada de la información presentada en gráficas y la elección y elaboración de gráficas para mostrar información contenida en los datos no son habilidades que se adquieran espontáneamente, sino que es necesario que sean objeto de la enseñanza.

ESTUDIOS SOBRE GRÁFICAS ESTADÍSTICAS

Diferentes niveles de lectura de datos

Un componente importante en el que han avanzado las investigaciones en didáctica de la estadística es en la identificación de diferentes niveles de lectura de los datos representados en tablas o gráficas. Por ejemplo, Curcio (1987) describe cuatro niveles distintos de lectura, de los cuales cada uno requiere mayor comprensión que sus antecesores.

1. El primer nivel es el de *leer los datos*: el lector logra identificar los datos individuales del conjunto y con ello puede responder las siguientes dos preguntas: ¿El dato x pertenece a la gráfica? ¿Cuál es la frecuencia del dato x?
2. El segundo nivel es el de *leer entre los datos*: el lector percibe relaciones relevantes entre los datos que le permiten responder preguntas como: ¿Cuál es el valor mínimo? ¿Cuál es el máximo? ¿Qué dato tiene mayor frecuencia? Este nivel incluye comparar e incluso identificar en la gráfica valores como la moda, la mediana y la media. Aunque de mayor dificultad, la estimación de las medidas de dispersión también están en este nivel.
3. El tercer nivel es el de *leer más allá de los datos*: el lector realiza predicciones e inferencias a partir de los datos sobre

información que no se reflejan directamente en la gráfica. Por ejemplo, ¿qué forma, media y/o dispersión tendrá la distribución de la población? Si se obtiene un nuevo dato, ¿es razonable pensar que pertenece al mismo conjunto (distribución)?

4. El cuarto nivel es el de *leer detrás de los datos*: un lector en este nivel, además de hacer inferencias o predicciones, también puede valorar la calidad de una gráfica y la validez y fiabilidad de los datos que representa. ¿La fuente es confiable? ¿Se pueden verificar las características del muestreo? ¿El muestreo fue apropiado para el tipo de estudio?

El aprendizaje de la lectura de gráficas se debe distribuir a lo largo de la educación preuniversitaria, y en el bachillerato el objetivo debiera ser que los estudiantes alcancen los niveles 3 y 4.

Dificultades y errores comunes en la elaboración de gráficas

a) Dificultades en la formación de una distribución de frecuencias

Los resultados de muchas investigaciones en las que se observa una variable cuantitativa de alguna población, en primera instancia, se representan en una lista de datos. Para graficarlos es necesario formar una distribución con ellos, es decir, identificar cada valor y su frecuencia. Una vez hecho lo anterior, se representan en el eje horizontal los diferentes valores de la variable y en el eje vertical sus frecuencias respectivas. Esta manera de proceder no es fácil de realizar, como se muestra en Arteaga, Batanero, Ortiz y Contreras (2011), quienes hallaron que más de 25% de una población de 207 futuros profesores fue incapaz de elaborar la distribución y graficarla.

Un error común que refleja la dificultad para formar la distribución de una lista de datos es la de considerar la pareja (orden de aparición, magnitud del número). Por ejemplo, se asigna la siguiente tarea:

Tira 10 veces un par de dados y suma los resultados. Anótalos en una lista y grafícalos.

Los estudiantes que ignoran cómo formar la distribución, ante un conjunto de datos como el que se muestra, hacen la gráfica de la figura 11a, en lugar de la gráfica de la figura 11b.

Por ejemplo, si se obtiene la lista de datos: 3, 5, 7, 6, 4, 7, 8, 3, 5,5. La gráfica de la izquierda no es correcta, pero es común. La gráfica de la derecha es la gráfica de la distribución.

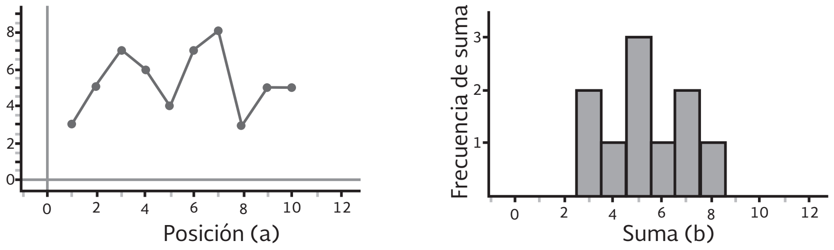


Figura 12. (a) Ejemplo incorrecto para graficar datos; (b) Forma correcta de graficar los datos del ejemplo.

b) Dificultades para definir las escalas o divisiones de los ejes

Los errores relacionados con las escalas, aunque pueden provenir del desconocimiento de convenciones acerca de la elaboración de las gráficas, también provienen de deficiencias en los antecedentes matemáticos que se requieren. Li y Shen (1992) encontraron que varios de sus estudiantes elegían una escala inadecuada para mostrar aspectos relevantes del conjunto de datos, o bien, que no alcanzaba a cubrir el campo de variación de la variable representada. Otros omitían las escalas en alguno de los ejes, horizontal o vertical (o en ambos), y muchos no especificaban el origen de las coordenadas. Hay otros que utilizan escalas no homogéneas.

Detrás de algunas de las anteriores dificultades se encuentra la tendencia a no ver en el eje X una recta numérica, sino sólo un espacio para indicar los valores que se presentan en los datos, como si fueran etiquetas.

ESTUDIOS SOBRE MEDIDAS DE TENDENCIA CENTRAL

Se suelen llamar medidas de tendencia central a la moda, la media y la mediana. Las definiciones matemáticas de estos conceptos se dieron en el capítulo 1 de este libro. Ahora revisaremos algunos aspectos que han surgido en las investigaciones sobre este tema. La mayoría de los estudios didácticos realizados a la fecha se enfocan en la noción de media y, por esta razón, en esta sección se comenta más extensamente este concepto que los otros dos.

La media aritmética de un conjunto de datos numéricos es un concepto engañoso, pues su definición es muy simple: la suma de datos entre el número de datos del conjunto. Aunque en algunas situaciones los estudiantes pueden tener algunas dificultades en cálculos relacionados con la media; por ejemplo, para calcularla a partir de datos agrupados o en problemas que suelen llamarse de media ponderada, en realidad, las verdaderas dificultades son conceptuales; es decir, cuando se trata de interpretarla, de atribuirle un significado apropiado, o de utilizarla en las diversas situaciones en las que su uso es pertinente.

Batanero (2000) menciona cuatro significados que suelen asociarse a la media aritmética, pero que no suelen ser bien entendidos por los estudiantes, de manera que no los utilizan en las situaciones en las que podrían ser útiles. Tales significados se agrupan en los siguientes cuatro rubros:

- Distribución equitativa de cantidades dispares.
- Estimación de una cantidad desconocida en presencia de errores de medida.
- El valor con mayor probabilidad al tomar un elemento al azar de una población.
- Representante de un conjunto de datos.

A continuación se precisa en qué consiste cada uno de ellos a través de ejemplos.

Distribución equitativa de cantidades dispares

En muchas situaciones en las que se tienen datos que varían uno del otro, se desea simplificar la situación pensando qué pasaría si no hubiera dicha variación; éste es un mecanismo del pensamiento que lo potencia. Por ejemplo, sabemos que la producción de petróleo en el país es variable; no obstante, se ofrecen noticias como la siguiente:

Produce Pemex 2 millones 562 mil barriles de petróleo diarios en enero. 01 de marzo de 2013. Boletín No. 28

Se entiende que algunos días de enero se producirán más de 2.56 millones, otros menos; de hecho, quizá los domingos no se produzca ninguno. Pero lo que uno entiende es que los 2.56 millones son un promedio. Este valor es accesible al pensamiento y a la memoria y es fácilmente comunicable, comparado con una tabla formada por la lista de la producción de cada uno de los 30 días de enero.

A pesar de que este significado es muy utilizado en la práctica y pareciera que no se presta a confusión, se ha encontrado que en ciertos contextos los estudiantes no lo interpretan adecuadamente. Por ejemplo, se les preguntó a estudiantes de secundaria lo siguiente:

Un estudio encontró que los estudiantes de escuelas primarias ven en promedio 3 horas diarias de televisión.

a) ¿Qué significa en promedio “3 horas de TV diarias” en este enunciado?

b) ¿cómo crees que se obtuvo este promedio?

Algunas de las respuestas al inciso a, son:

- Los niños ven la TV aproximadamente tres horas diarias.
- La mayoría de niños ve la TV tres horas diarias.

Ninguna de ellas es adecuada; la primera porque podría haber algunos niños que vean la TV cero horas, otros 10 horas diarias y el promedio podría ser 3; en este caso no se puede decir que 0 y 10 son aproximadamente 3. La segunda respuesta tampoco es adecuada porque podría no haber ningún niño que vea la TV las 3 horas diarias y, no obstante, obtenerse el promedio de 3.

Con referencia al inciso b de la situación planteada arriba, también hay respuestas que no coinciden con lo que supone es la normatividad:

- Se observó que un niño veía la TV tres horas diarias.
- Se observaron muchos niños y se vio que veían la TV alrededor de 3 horas.
- Se vio que la mayoría de los niños veían la TV 3 horas diarias.

Las anteriores comprensiones inadecuadas sobre la media provienen del uso informal de este concepto sin hacer las precisiones necesarias relacionadas con la definición técnica y sus propiedades.

Otras interpretaciones encontradas en niños más pequeños para la media se presentan en Watson y Moritz (2000). Estos autores estudian la forma en que los niños de diferentes edades comprenden el concepto de "promedio" y el uso que hacen de éste en situaciones de la vida diaria. Realizaron entrevistas con alumnos de tercero de secundaria en las que plantean preguntas como las siguientes: “¿Has escuchado la palabra promedio? ¿Qué significa para ti?” “En promedio, las familias australianas tienen 2.3 niños. Explica qué significa para ti esta frase”. “Supongamos que 10 familias tienen en promedio 2.3 niños. Entre ellas, la familia Grant tiene 4 y la familia Cooper tiene 1, ¿cuántos hijos podrían tener las restantes 8 familias?”. Con base en las respuestas recibidas, identificaron seis niveles de comprensión de los promedios:

1. Nivel *pre-promedio*: los niños no usan ningún término de promedio y sólo imaginan historias con cierta relación con el contexto de la pregunta.

2. *Uso coloquial de un promedio* utilizan términos como "es normal o es okey". Imaginan ideas relacionadas al contexto para apoyar sus respuestas; por ejemplo, en relación con la pregunta responden historias como "*Que tienen dos niños grandes y otro que no ha crecido todavía*", "*el .2 es un niño que tiene 2 años ahora y cuando cumpla 10, contará como 1 y entonces el número promedio de niños será 3*".
3. *Estructura múltiple del promedio*, manejan dos o más ideas como el *más común*, el de en medio y, en ocasiones, el algoritmo al describir un promedio.
4. *Representación con promedio* se refieren directamente al algoritmo de la media en relación con las situaciones. Además son conscientes de que la forma decimal de una media es consecuencia del algoritmo y expresan alguna idea sobre la naturaleza de la media como representante de un conjunto.
5. Los estudiantes además son capaces de aplicar su conocimiento de la media a situaciones complejas, como completar un conjunto de valores para obtener una media dada o calcular una media ponderada, pero no ambas.
6. Hacen todo lo anterior incluyendo las dos últimas tareas mencionadas.

Varias de las dificultades que se observan en los estudios con niños pequeños se vuelven a encontrar ligeramente distintos en estudiantes de bachillerato e incluso universitarios. Por ejemplo, Mayen (2009) hizo un estudio con alumnos de secundaria y bachillerato en México tomando preguntas de investigaciones anteriores. Sólo 52% de los estudiantes de secundaria y 80% de los de bachillerato respondió correctamente a la siguiente situación "Un periódico dice que 2.2 es el número medio de hijos por familia en México. Explica qué significa para ti esta frase". Al pedirles un ejemplo concreto de distribución de 10 familias para obtener un número medio de 2.2 hijos, el número de respuestas correctas en estudiantes de bachillerato bajó a 56%.

Aunque la interpretación de la media como distribución equitativa de cantidades dispares es la más cercana al algoritmo, en el sentido de que se deduce de las operaciones aritméticas que se realizan para calcularla, no está exenta de dificultades. La siguiente interpretación presenta mayor grado de dificultad.

Estimación de una cantidad desconocida, en presencia de errores de medida.

En la teoría de errores de medidas físicas la media juega un papel importante y de gran utilidad. En efecto, cuando en un trabajo experimental se realizan varias mediciones de una magnitud física, en general, no se obtiene el mismo resultado. Surge entonces el problema de cuál es la mejor aproximación a la verdadera medida de la magnitud en cuestión. La teoría de los errores establece que la mejor aproximación es la media aritmética cuando las mediciones cumplen las siguientes características:

- Cada medición es independiente de cualquier otra.
- Se pueden cometer errores con la misma frecuencia, tanto por defecto como por exceso.

La media aritmética se aproximará tanto más al valor verdadero de la magnitud cuanto mayor sea el número de medidas, ya que globalmente los errores aleatorios se compensan. Este uso de la media de un conjunto de medidas de un objeto como la mejor estimación de la medida real del objeto no es una idea que surja espontáneamente en los estudiantes, y aunque el profesor lo mencione en su clase, ellos suelen no utilizarla para resolver problemas relacionados. Para comprobarlo, basta hacer preguntas como la siguiente:

> Un objeto pequeño fue pesado separadamente por nueve estudiantes en una clase de ciencias, utilizando la misma escala. Los pesos (en gramos) anotados por cada uno de ellos se muestran a continuación.
>
> 6.1 6.0 6.0 5.9 6.4 6.3 6.0 6.2
>
> Los estudiantes quieren determinar tan acertadamente como sea posible el peso real de este objeto. ¿Qué valor les propondrías?

Muchos estudiantes tenderán a elegir 6.0 “porque es el que más se repite”, otros dirán que cualquiera podría servir, algunos proponen el punto medio del rango; muy pocos propondrán la media, que de acuerdo con la teoría de errores sería la mejor estimación. Si los alumnos son los que obtienen los pesos, optan por el valor de “quien tomó el peso con más cuidado”, lo que hace intervenir apreciaciones subjetivas.

Como se verá más adelante, los problemas significativos en los que se debe aplicar la media se dan en contextos que requieren de otras consideraciones; por ejemplo, de la dispersión de los datos. Con base en ellas el analista no aplica rígidamente la regla; por ejemplo, la siguiente modificación transforma el problema anterior:

> Un objeto pequeño fue pesado separadamente por nueve estudiantes en una clase de ciencias, utilizando la misma escala. Los pesos (en gramos) anotados por cada uno de ellos se muestran a continuación.
>
> 6.1 6.0 6.0 15.9 6.4 6.3 6.3 6.0 6.2
>
> Los estudiantes quieren determinar tan acertadamente como sea posible el peso real de este objeto. ¿Qué valor les propondrías?

La media en este problema es: 7.24, que sería muy probablemente una mala estimación ocasionada por el valor atípico 15.9. El analista supone que este valor es espurio, debido a un error diferente a los errores de medida esperados; entonces puede optar por dos posibilidades: 1) Excluir el valor atípico y obtener la media del resto de los datos: 6.13; 2) elegir la mediana: 6.2.

La amplia aplicación de la estadística al análisis de diferentes tipos de medidas experimentales es uno de los motivos por los que se imparten cursos de Estadística en las carreras científicas. Para explotar su potencial, el estudiante debe aprender a interpretar los diferentes conceptos y procedimientos estadísticos de manera adecuada y flexible; uno de éstos es la media, que aunque parezca muy simple, encierra ideas profundas que el estudiante debe asimilar. La siguiente interpretación es otro ejemplo.

El valor con mayor probabilidad al tomar un elemento al azar de una población.

La relación de la media con los fenómenos aleatorios a veces no es tan clara porque la media se presenta en estadística descriptiva de una manera y en probabilidad de otra, a saber, como el valor esperado. Para ver dicha relación conviene referir la situación de muestreo simple que menciona Batanero (2000), en la que se eligen elementos de una población numérica cuya media es conocida:

> La altura media de los alumnos de un colegio es 1.40 m. Si extraemos una muestra aleatoria de 5 estudiantes y resulta que la altura de los 4 primeros es de 1.38 m, 1.42 m, 1.60 m y 1.40 m. ¿Cuál sería la altura más probable del quinto estudiante? (Batanero, 2000).

El error más frecuente en el problema anterior consiste en buscar el valor para el que la muestra tenga una media de 1.40 m, en este caso: 1.20 m. Éste surge de la creencia de los estudiantes de que la solución debe encontrarse mediante un procedimiento matemático preciso; la solución correcta, que es 1.40m, no se obtiene mediante procedimiento alguno, si no que se basa en la comprensión de lo que es una media en una distribución.

Como hemos mencionado, las interpretaciones dependen de condiciones particulares del contexto. En el caso anterior, se utiliza el hecho (conocido, pero no explícito en el enunciado) de que la distribución de las alturas tiene una forma aproximadamente normal (en forma de campana y simétrica respecto a su media), de manera que la media coincide con el valor más probable.

Batanero (2000), al destacar este significado o interpretación de la media, ofrece algunos ejemplos comunes en estadística:

> Por ejemplo, al predecir la esperanza de vida o el beneficio esperado en una inversión en bolsa, se toma la media de la variable en la población como predicción, como valor esperado, por sus propiedades muestrales derivadas del teorema central del límite. Del concepto de valor esperado se derivan muchos modelos de predicción, como los distintos tipos de regresión. Así, cuando predecimos el peso de una persona en función de su altura, usamos el peso promedio de todas las personas que en la población tienen la altura (p. 4).

Estos comentarios son válidos, pues las variables de la que provienen los datos tienen distribuciones normales.

Representante de un conjunto de datos

Este significado de la media se utiliza muy a menudo, pero es de gran dificultad para los estudiantes. El ejemplo clásico es el promedio de calificaciones como representante o resumen de las diferentes calificaciones obtenidas por un estudiante en las asignaturas del programa escolar. En este contexto familiar todo mundo entiende la función de la media (o promedio) como representante del conjunto de calificaciones, pero cuando se formulan problemas en otros contextos, esta interpretación no surge de manera espontánea y natural. Por ejemplo, en el siguiente problema de comparación de datos:

1. En la tabla de abajo se presentan las estaturas en metros de 10 hombres holandeses:

1.87	1.71	1.75	1.83	1.85	1.92	1.92	1.86	1.70	1.88

a) ¿Cuál es la estatura de los hombres en Holanda?

b) ¿Qué valor o número consideraste como representante de las alturas de los holandeses?

2. En la tabla de abajo se presentan las estaturas en metros de ocho hombres mexicanos:

1.74	1.71	1.96	1.80	1.76	1.72	1.68	1.62

a) ¿Cuál es la estatura de los hombres en México?

b) ¿Qué valor o número consideraste como representante de las alturas de los mexicanos?

3. ¿Cuántos centímetros son más altos los holandeses que los mexicanos?________________________________

4. Reúnanse con otros compañeros y comenten sus respuestas. Cada quien argumente sus propuestas y lleguen a un acuerdo sobre las respuestas correctas. En caso de duda consulten con su profesor (Sánchez, *et al.,* 2012).

En un estudio de Orta (2009) con estudiantes de secundaria, encontró que responden espontáneamente de formas muy variadas a las preguntas, pero sólo 5% utiliza la media aritmética en sus respuestas.

Los problemas de comparación de grupos de datos o de distribuciones son muy apropiados para que los estudiantes aprendan a utilizar la media como un recurso para resolverlos y, con ello, incorporar dicho significado a su formación estadística. En la siguiente sección, referente a la dispersión de los datos, volveremos a este punto.

ESTUDIOS SOBRE DISPERSIÓN

Hace relativamente poco tiempo, Shaughnessy (1997) señaló que hasta esa fecha no había habido investigaciones didácticas que estudiaran las dificultades, concepciones, problemas y formas de enseñar el concepto de variación estadística. Mostró cómo los investigadores en educación estadística habían estado más interesados en el concepto de media. Esta escasa o nula investigación sobre la variación, que se puso de relieve en ese entonces, contrasta con la opinión general de los estadísticos acerca de que la variación es el corazón de la estadística (Moore; 1998). A partir de entonces se han hecho importantes estudios que plantean preguntas acerca del aprendizaje de la variación.

La *dispersión* es un sinónimo de *variación* cuando ésta se refiere a un conjunto de datos numéricos. Las medidas de dispersión más comunes son el rango, la desviación media, la desviación estándar y los cuartiles (o más en general, los percentiles). En la enseñanza el tratamiento de estos conceptos ha consistido, al igual que con el concepto de media, en buscar que los estudiantes se aprendan la fórmula para calcularlas, bajo el supuesto de que esto puede permitir su interpretación y uso en las situaciones- problema en las que este concepto es útil. Nada más alejado de la realidad.

A continuación se presentan algunas ideas básicas sobre el riesgo basadas en la exposición de Fischhoff y Kadvany (2011).

En particular, para los fines de este libro basta mencionar algunas ideas de tres temas que abordan estos autores: la definición de riesgo, el análisis del riesgo y la toma de decisiones en situaciones de riesgo. El riesgo se presenta cuando hay potenciales resultados no deseados que pueden traer como consecuencia pérdidas o daños. Definir el riesgo significa especificar los resultados valiosos y los no deseados en un orden que refleje el valor que se les atribuye. En muchas ocasiones, se pueden medir los resultados no deseados o valiosos; por ejemplo, las tasas de mortalidad o el producto interno bruto. También hay eventos en los que no hay acuerdo de cómo medirlos: pobreza, bienestar o sustentabilidad. Finalmente, hay algunos en que la sola idea de medirlos es controversial, así son las amenazas a la justicia o los daños ecológicos a la naturaleza.

Una vez que se identifican los riesgos, el problema es determinar qué tan grandes o prejudiciales son y cuáles son las posibles causas que los gobiernan. El análisis del riesgo generalmente son constructos intrincados, a menudo integran los análisis de científicos de diferentes áreas del conocimiento y toman en cuenta una diversidad de fuentes de datos o información. No obstante, hay una variedad tan amplia de fenómenos, que la lógica del análisis del riesgo puede consistir en simples conteos de datos, con cuyos resultados se puedan llevar a cabo conocidos análisis estadísticos, hasta complejas simulaciones de modelos cibernéticos que contengan información científica de áreas especializadas además, de procesos estocásticos complejos.

El análisis del riesgo ofrece información para la toma de decisiones. La teoría de la toma de decisiones en situaciones de riesgo tiene dos aspectos; por un lado, define reglas abstractas sobre lo que debería hacer la gente en situaciones de riesgo; por otro, estudia lo que hace la gente cuando se enfrenta realmente a tales situaciones: "Si la gente no sigue las reglas, quiere decir que o las personas necesitan ayuda o las reglas necesitan una revisión" (Fischhoff y Kadvany, 2011, p. 65).

En los casos en que el análisis del riesgo ofrece un conjunto ordenado de posibles resultados, la regla para tomar decisiones es simple: "elija la opción cuyo resultado produzca algo con la mayor cantidad del valor que usted desea tener (dinero, descanso, acres de tierras húmedas, etc.)" (Fischhoff y Kadvany, 2011, p. 65). La ordenación de los resultados generalmente contiene incertidumbre y está asociada a distribuciones de probabilidad. Entonces la regla debe combinar el valor deseado que proporciona cada resultado con la probabilidad de que ocurra para cada posible decisión.

Un problema en el contexto de juegos

No iremos más allá de los comentarios anteriores, pues son suficientes para indicar la naturaleza del contexto de riesgo en el problema que se propone a continuación. Al analizar el problema de debe aclarar el papel de la media y la dispersión para la toma de decisiones en la situación de riesgo que se propone.

En una feria se invita a los asistentes a participar en uno de dos juegos. Juan puede participar en un juego, pero no en ambos. Para saber por cuál decidirse observa, anota y ordena los resultados de dos muestras de 10 personas que han participado en cada juego. Las pérdidas (-) o premios (+) en efectivo que han obtenido las 20 personas se muestran en las siguientes listas:

Juego 1

15	-21	-4	50	-2	11	13	-25	16	-4

Juego 2

120	-120	60	-24	-21	133	-81	96	-132	18

Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías? Explica las razones de tu respuesta.

Una respuesta relativamente aceptable es que es igual jugar en cualquiera de los dos juegos, esto con base en el argumento de que ambos conjuntos tienen la misma media (μ = 4.9), es decir, a la larga en cualquiera de los juegos se ganará en promedio 4.9 unidades. No obstante, una respuesta más avanzada debe tener en cuenta la dispersión. Es notorio que en los datos del segundo juego hay mucha más dispersión que en los del primero (figura 13). En este contexto, la dispersión se asocia al riesgo: el juego 2 es más arriesgado, pues aunque con base en los datos se puede conjeturar que es posible ganar cantidades mayores que en el juego 1 (el máximo en el juego 2 es 133, mientras que en el juego 1 es 50) también es posible perder más (el mínimo en el juego 2 es -132, mientras que en el juego 1 es -25). La decisión depende de las preferencias y condiciones económicas del jugador, si es propenso al riesgo y tiene suficientes recursos para jugar, podría optar por el juego 2; si es conservador o no tiene suficientes recursos, le conviene elegir el juego 1.

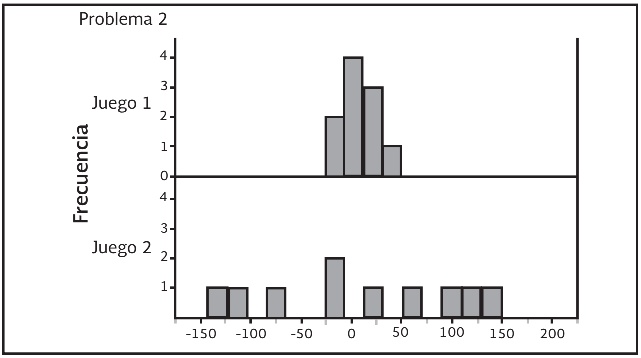


Figura 13. Gráfica del problema de riesgo.

Análisis del problema. La idea del problema es propiciar que el estudiante se ubique en una situación de riesgo, de manera que se dé cuenta que la elección de un juego u otro conlleva consecuencias en términos de beneficios-pérdidas. El primer paso en el análisis es preguntarse en qué juego se gana más dinero globalmente; esta pregunta se la hacen los estudiantes de manera natural, pero la respuesta que ofrecen es muy variada. Nuevamente pensar en la ganancia promedio de cada juego (media aritmética de los datos) y compararlas es una idea que pocos, si alguno, la consideran; la mayoría omiten este paso que es crucial para el análisis. Sus comparaciones suelen ser resultado de una impresión de riesgo que les producen los datos individuales de la distribución: unos comparan los máximos de las distribuciones y hacen afirmaciones como “en el juego 2 se gana más”, otros comparan los mínimos “en el juego 1 se pierde menos”; otros podrían ignorar la cantidad precisa que se gana o se pierde y sólo cuentan cuántos perdedores (cantidades negativas) y cuántos ganadores (cantidades positivas) tiene cada juego, decidiendo que da lo mismo escoger uno u otro.

Una vez que se determina con base en la ganancia promedio que los dos juegos son equivalentes, se debe formular la pregunta ¿en cuál juego se corre más riesgo? Inferir que el juego con mayor dispersión es más arriesgado que el juego con menor dispersión ofrece un significado pertinente a la teoría del riesgo. El problema de valorar el riesgo consiste en ponderar que mientras en el juego con mayor dispersión se puede tener la fortuna de ganar más que en el otro, pero también se corre el riesgo de perder más. En el juego con menor riesgo se puede ganar menos que en el primero, pero también se puede perder menos. No hay una regla general que indique cuál es la mejor decisión, pues depende de la situación económica particular del jugador, así como de sus creencias y actitudes ante el riesgo.

Una dificultad que suelen encontrar los estudiantes surge de la falta de información acerca de la naturaleza del juego y sobre la manera particular en que un jugador gana determinada cantidad. Muchos estudiantes llenan este vacío imaginando situaciones: unos pueden pensar en una urna con bolas numeradas, otros en juegos de caballos, etc. Encontramos que algunos estudiantes imaginan que el primer dato del primer juego (15) corresponde a un jugador y que el primer dato del segundo juego corresponde a su oponente (120) y concluyen que el jugador del segundo juego le ganó al primero (por 105). Situaciones en las que se pueden coleccionar datos ignorando el modelo subyacente son frecuentes en la investigación estadística. La idea del problema es propiciar que el estudiante realice un estricto análisis de los datos interpretando la situación a pesar de dichas restricciones. El siguiente ejemplo es también de una situación de riesgo y en él es más clara la ausencia de un modelo para explicar el comportamiento de los datos.

Un problema en el contexto de salud-enfermedad

Ejemplo de problema es situaciones de riesgo.

> Considera que debes aconsejar a una persona que padece una enfermedad grave, incurable y mortal, pero que es tratable con medicamentos que pueden extender la vida por varios años más. Es posible elegir entre tres tratamientos. Las personas tienen diferentes reacciones a las medicinas, para algunas tienen el resultado previsto, mientras que para otras pueden ser más benéficas o más prejudiciales. En las siguientes listas se muestran los años que han vivido varios pacientes que se han tratado con una de las opciones mencionadas; cada dato de las listas corresponde al tiempo que ha sobrevivido un paciente con el respectivo tratamiento. Después se muestran las gráficas correspondientes a los tratamientos.
>
> Tiempo en años (tratamiento 1)
>
> 5.2 5.6 6.5 6.5 7.0 7.0 7.0 7.8 8.7 9.1
>
> Tiempo en años (tratamiento 2)
>
> 6.8 6.9 6.9 7.0 7.0 7.0 7.1 7.1 7.2 7.4
>
> Tiempo en años (tratamiento 3)
>
> 6.8 6.8 6.9 7.0 7.0 7.1 7.1 7.1 7.2 7.4
>
> ¿Que tratamiento recomendarías? ¿Por qué?

El primer paso en el análisis de este problema, al igual que los anteriores, es calcular el tiempo medio de vida para cada tratamiento ($\mu = 7.04$). Con base en esta información se podría decir que cualquier tratamiento es igual; no obstante, en la medida en que lo que está en juego es de suma importancia, parece imprescindible hacer un análisis (de riesgo) más detenido. Claramente el tratamiento 1 tiene mayor dispersión (figura 14) y, por lo tanto, ofrece más riesgo, en el sentido mismo del problema anterior: los datos indican que se puede vivir más que el tiempo que indican los datos en cualquiera de los otros dos tratamientos ($9.1 > 7.4$), pero también se puede vivir menos ($5.2 < 6.8$). ¿Cuál es la decisión adecuada entre elegir el primer tratamiento y cualquiera de los otros dos? Nuevamente, no hay una regla general que prescriba la mejor elección, aunque en este caso una decisión conservadora podría ser más popular que una decisión arriesgada. Decidir entre el tratamiento 2 y el 3 es más difícil, pero si el criterio de menor riesgo se ha tomado y éste se asocia con la desviación estándar muestral, los datos del tratamiento 2 indican menor dispersión (S_2 = 0.17, S_3 = 0.18), aunque la diferencia es tan pequeña que se pueden considerar equivalentes.

Análisis del problema. El análisis de este problema es muy similar al anterior sobre la elección entre dos juegos. El primer paso es comparar las medias, después considerar con cuidado la dispersión. Como se ha indicado, este sencillo procedimiento no suele estar dentro de las estrategias espontáneas de los estudiantes y bien puede ocurrir que hayan entendido que en el contexto de pérdida-ganancia es conveniente comparar las medias de los conjuntos de datos y no transferirlo a las situaciones de tiempo de vida en contextos de salud-enfermedad. Con relación al análisis de la dispersión, es frecuente que un mismo alumno sea propenso al riesgo en el caso de juegos de pérdida-ganancia de dinero, y sea adverso al riesgo en el caso de salud-enfermedad o viceversa. Lo importante de la consideración del riesgo es que lo interpreten de manera adecuada y no caigan en el error de creer que "mayor dispersión significa mayor tiempo de vida", pues esto es falso.

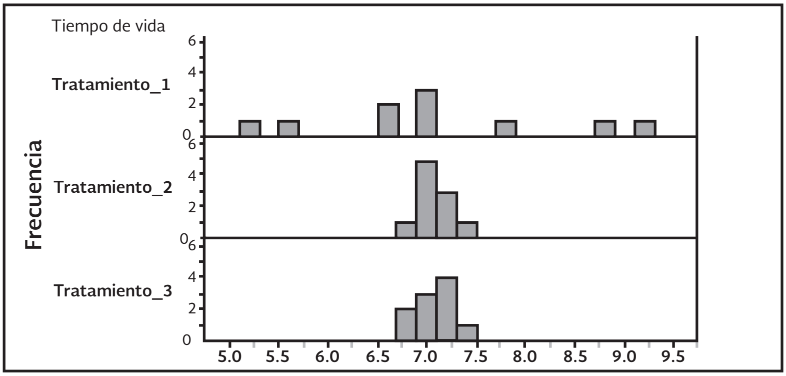
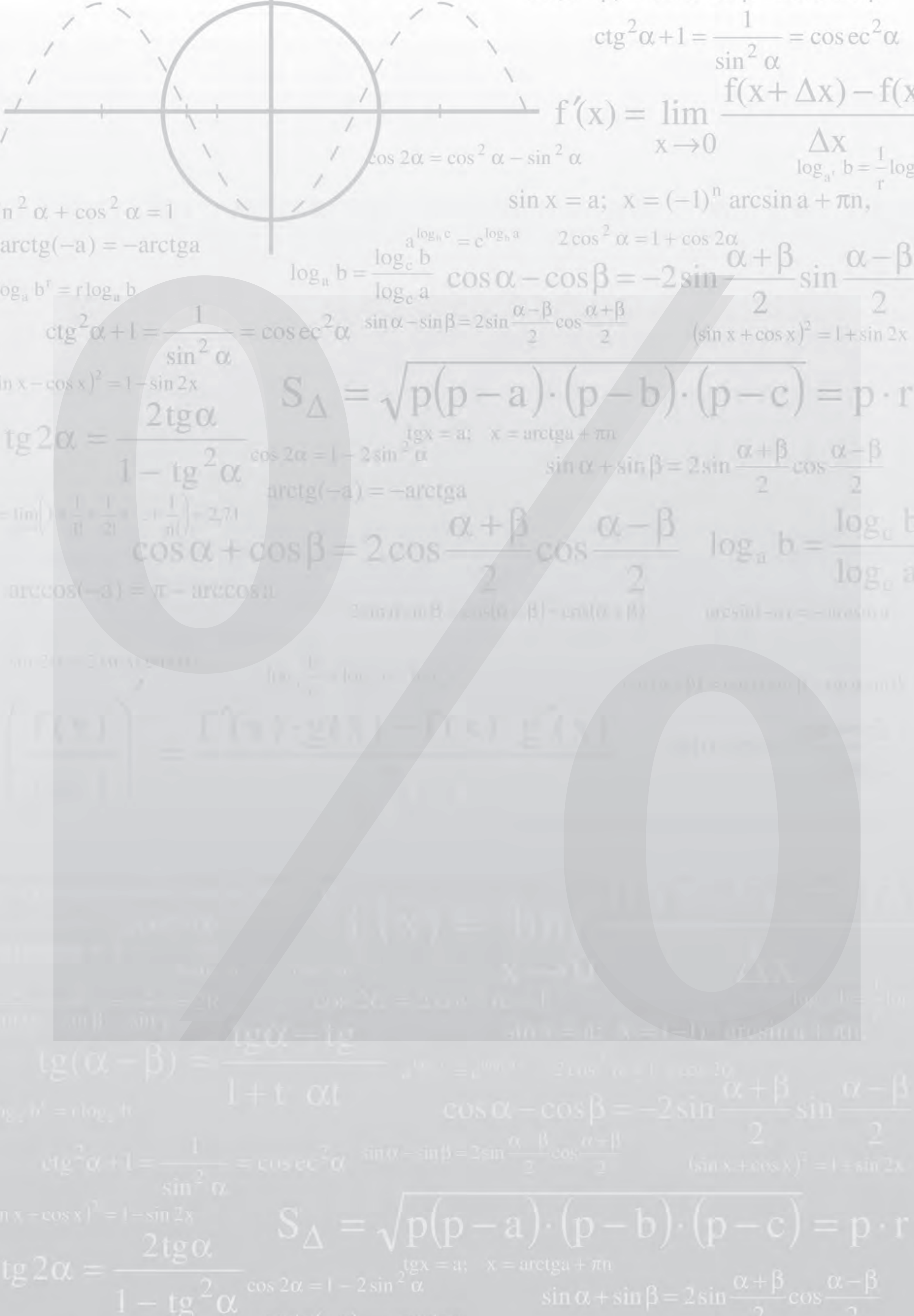


Figura 14. Gráfica de los datos del problema del tiempo de vida.

Conclusión. Se han propuesto tres problemas que pueden ser útiles para diseñar lecciones de estadística para desarrollar el razonamiento estadístico de los estudiantes en los temas de media y dispersión. Las razones para hacerlo son variadas: la importancia de que el profesor cuente con problemas para cubrir los contenidos curriculares, las dificultades detectadas en el tema en cuestión y la necesidad de darles un significado que sea accesible a los alumnos.

Los problemas se han adaptado al formato de comparación de conjuntos de datos, pues prefiguran problemas de inferencia, atraen el interés de los estudiantes, no son triviales y su solución requiere de los instrumentos elementales de la estadística. También se han inscrito en los contextos de mediciones repetidas (ya señalado por Batanero, 2000) y de riesgo, pues ambos ofrecen oportunidades para enriquecer el significado de las nociones de media y dispersión. La intención es explorar recursos que permitan ir más allá de las interpretaciones de la media y la dispersión que se constriñen a contextos puramente matemáticos. Se han hecho algunas recomendaciones para implementarlos en clase, acordes con el espíritu actual de las recomendaciones curriculares: dejar a los estudiantes que propongan soluciones, promover la discusión y prever los errores y falsas concepciones que ya han sido detectadas en la investigación. En el futuro, investigaciones empíricas y prácticas escolares podrían proporcionar evidencias de las ventajas y desventajas del uso de este tipo de problemas en clase.



ctg²α + 1 = 1/sin²α = cosec²α
f′(x) = lim x→0 (f(x+Δx) − f(x))/Δx
cos 2α = cos²α − sin²α
sin x = a; x = (−1)ⁿ arcsin a + πn,
sin²α + cos²α = 1
arctg(−a) = −arctga
2cos²α = 1 + cos 2α
log_a b = log_c b / log_c a
cos α − cos β = −2 sin (α+β)/2 sin (α−β)/2
sin α − sin β = 2 sin (α−β)/2 cos (α+β)/2
(sin x + cos x)² = 1 + sin 2x
(sin x − cos x)² = 1 − sin 2x
S_Δ = √(p(p−a)·(p−b)·(p−c)) = p·r
tg 2α = 2tgα / (1 − tg²α)
tgx = a; x = arctga + πn
cos 2α = 1 − 2sin²α
sin α + sin β = 2 sin (α+β)/2 cos (α−β)/2
cos α + cos β = 2 cos (α+β)/2 cos (α−β)/2
arccos(−a) = π − arccos a

3. ESTRATEGIAS PARA LA ENSEÑANZA DE LA ESTADÍSTICA

La investigación en educación estadística ha desarrollado propuestas para transformar la enseñanza de la materia y ofrecer mayores probabilidades de que los estudiantes desarrollen competencias estadísticas. En este capítulo se expondrán tales estrategias y los argumentos que las justifican.

EL CONCEPTO DE COMPETENCIAS

Existe un descontento generalizado por los logros que se obtienen de la educación matemática. Y no sólo porque haya bajos índices de aprobación en la materia, sino porque aun los estudiantes que acreditan con altas notas los cursos de matemáticas no saben utilizar adecuadamente sus conocimientos en los contextos de su vida personal o profesional donde las matemáticas serían útiles. En estos casos se habla de un conocimiento enciclopédico y no un conocimiento funcional.

En estadística, por ejemplo, hay quienes aprenden a leer y construir gráficas en el salón de clase, pero no saben ni quieren leer o interpretar las gráficas periodísticas o del INEGI; tampoco se les ocurre que una gráfica podría contribuir al análisis y comunicación de resultados en sus actividades profesionales.

¿Qué son las competencias?

Conviene revisar algunas definiciones del concepto *competencia* para formarse una idea del núcleo común que comparten. Para Perrenoud (2002), "una competencia es una capacidad para movilizar diversos recursos cognitivos para hacer frente a un tipo de situaciones". Las personas tienen recursos cognitivos variados, se dice que son competentes si saben ponerlos en acción y los movilizan para realizar las tareas derivadas de las situaciones que enfrentan. Una definición más conocida es la que se formula en el proyecto Tunning (Beneitone *et al.*, 2007): "El concepto de competencia se entiende como una combinación dinámica de atributos, en relación con conocimientos, habilidades, actitudes y responsabilidades, que describen los resultados de los aprendizajes de un programa educativo o lo que los estudiantes son capaces de demostrar al final del proceso educativo".

Las competencias se ubican en la relación entre conocimiento y acción, entre teoría y práctica. De lo que se trata es de que en la educación se asuma el compromiso de formar ciudadanos capaces de utilizar sus conocimientos en la solución de los problemas que encontrarán en su vida personal, social y laboral. Aunque nunca se ha dicho que la educación persiga lo contrario, en realidad se ha asumido que la adquisición de conocimientos es una condición previa para su aplicación y que la escuela se encarga de proporcionarlos. Se supone que una base sólida en este sentido servirá para aplicar tales conocimientos en la vida personal, social y laboral/profesional. Pero estos supuestos, aparentemente válidos, los cuestiona una realidad en la que la gente aprende a actuar al margen de los conocimientos adquiridos. Peor aún, cuando intentan ponerlos en juego no saben cómo hacerlo o lo hacen de manera deficiente o equivocada.

La competencia matemática

Dado que hay una estrecha relación entre la matemática y la estadística, conviene entonces repasar algunas formulaciones que aclaran lo que significa la competencia matemática. En el proyecto PISA se define como "capacidad de un individuo para identificar y comprender el papel que las matemáticas juegan en el mundo, realizar razonamientos bien fundados y utilizar e involucrarse en las matemáticas de manera que satisfagan las necesidades de la vida del individuo como ciudadano constructivo, comprometido y reflexivo" (INCE 2004, p. 12) [PISA].

En esta declaración se subrayan los aspectos que diferencian a la propuesta de la educación por competencias de otras, como la de resolución de problemas o la guiada por el constructivismo. Esto es, se enfatiza en la relación de la matemática con el mundo (aplicaciones) y en su utilidad en la vida del individuo. Parece mentira que la sencillez y especificidad de esta declaración origine serias discusiones, ya que en ningún currículo o programa de estudios se ha pretendido que lo que se enseñe y aprenda el estudiante de las matemáticas en la escuela no le sirva mayormente su vida. Pero en los hechos, en muchos casos, la utilidad que obtienen en la enseñanza tradicional de las matemáticas se reduce a no reprobar la materia y a librar el obstáculo que representa para realizar el proyecto de obtener un título profesional; después de esto difícilmente se acuerdan de lo aprendido en esta materia.

Algunos investigadores que analizan y promueven el currículo por competencias consideran que cuando se habla de la utilidad de la matemática en la vida del individuo se corre el peligro de enfocarlo desde un punto de vista estrecho, que puede llevar a pensar en la "matemática de la tiendita", es decir, en la aritmética elemental que se necesita para saber cuánto se tiene que pagar en el mercado. Para evitar esta connotación o reduccionismo proponen el concepto de "contextos o situaciones socialmente relevantes" en las cuales el conocimiento matemático es significativo. Desarrollar una competencia matemática sería entonces ser capaz de utilizar el conocimiento matemático en las situaciones socialmente relevantes. Hay que notar, que para la matemática en general, no hay consenso acerca de cuáles son tales situaciones socialmente relevantes y que la elaboración de un currículo por competencias debería partir de su constitución. Pero el caso de la estadística es diferente, pues en esta disciplina están muy bien definidas, de tal manera que se puede hablar de competencias estadísticas.

Relaciones entre la matemática y la estadística

Para entender el papel de la estadística en el currículo por competencias, conviene tener en cuenta cuál es la discusión acerca de la relación entre la estadística y la matemática. Varios estadísticos y educadores estadísticos defienden que la estadística no es una rama de las matemáticas, ni siquiera dentro de las matemáticas aplicadas. La estadística tiene características propias que conviene tener en cuenta para entender mejor las propuestas de cómo enseñarla.

Entre las diferencias fundamentales se encuentran:

a) La consideración de los datos, los cuales en estadística, a diferencia de las matemáticas, no se pueden separar del contexto del que provienen.
b) La importancia de la recolección de datos, así como las circunstancias en las que se lleva a cabo.
c) La centralidad de la variabilidad.
d) La presencia de la incertidumbre en sus resultados.

Esta discusión sobre si la estadística es o no parte de las matemáticas se resuelve si se enfoca la enseñanza de las matemáticas desde un currículo por competencias. Pues se puede pensar que la estadística proporciona situaciones socialmente relevantes en las que se utilizan las matemáticas. Esto quiere decir, las situaciones de la estadística son apropiadas para desarrollar competencias matemáticas. De esta manera se debe considerar entonces que la noción de competencia estadística está relacionada con las competencias matemáticas.

El concepto de competencia estadística

Se define mediante los siguientes puntos:

a) Poseer y saber utilizar los conocimientos básicos de la estadística.

b) La habilidad para interpretar y evaluar críticamente la información estadística, argumentos relacionados con los datos o con fenómenos estocásticos.

c) La habilidad para discutir y comunicar sus reacciones a tal información estadística, tal como comprender el significado de la información y sus implicaciones.

d) La actitud de formular preguntas que puedan responderse mediante la recolección y análisis de datos.

e) Las destrezas para diseñar y llevar a cabo investigaciones estadísticas.

La estadística es naturalmente multidisciplinar, en el sentido de que su campo de acción se sobrepone con otras disciplinas. La definición de Moore (1998) es relevante en este sentido, pues “la estadística es un método intelectual general que se aplica dondequiera que haya datos, variación y azar. Es un método fundamental porque los datos, la variación y el azar son omnipresentes en la vida moderna” (p. 134). Para Cobb y Moore (2000), “la estadística es una disciplina metodológica. Ella existe no por sí misma sino para ofrecer a otros campos de estudio un conjunto coherente de ideas y herramientas para tratar con datos. La necesidad de una disciplina como ésta, surge de la omnipresencia de la variabilidad”. Sin duda, estas características de la estadística son favorables para el enfoque por competencias.

Un punto de referencia importante para definir las competencias acerca de cualquier disciplina es observar lo que hacen los expertos en esa disciplina. En el caso de la estadística, Wild y Pfannkuch (1999) hicieron una investigación sobre cómo piensan y actúan los estadísticos cuando trabajan en su profesión. De ese estudio formularon un modelo del pensamiento estadístico, que se expone en la siguiente sección.

El modelo de pensamiento estadístico de Wild y Pfannkuch (1999)

El modelo tiene cuatro dimensiones (ver figura 14), las cuales organizan el conjunto de aspectos que pone en juego un estadístico cuando resuelve problemas estadísticos. El primero es el ciclo investigativo (Dimensión 1), formado por los cinco pasos (PPDAC = Problema, Plan, Datos, Análisis, Conclusiones), problema que se ponen en juego al desarrollar uno: el interés y la habilidad de formular un problema susceptible de responderse con datos; el diseño de un *plan* para obtener los datos; la obtención o recogida de *datos* llevando a cabo dicho plan; el análisis de los datos y obtención de *conclusiones*. Para resolver un problema el estadístico profesional sabe que debe desarrollar un ciclo investigativo; cuando lo desarrolla pone en práctica ciertos esquemas de pensamiento que lo guían, los cuales constituyen los tipos fundamentales de pensamiento (Dimensión 2) y son:

a) *La necesidad de los datos*. El matemático enfrenta muchos de sus problemas mediante la reflexión y la representación de sus ideas en el papel; en contraste, el estadístico sabe que para resolver un problema requiere datos.

b) *La transnumeración*. Consiste en las diversas transformaciones a que se someten los datos, la realización de diversos tipos de gráficas y el cálculo de números resúmenes (media, moda o variancia; rango, desviación media, desviación estándar) para tratar de revelar la historia que cuentan los datos.

c) *Consideración de la variación*. La búsqueda de patrones de variación en los datos, así como su descripción, explicación o control, de acuerdo con el tipo de problema y el contexto al que pertenecen los datos.

d) *Razonamiento con modelos*. Consiste en aplicar las herramientas estadísticas al problema y situación de la que provienen los datos. Estas herramientas son las que se proporcionan en los cursos tradicionales de estadística: distribuciones, muestreo, intervalos de confianza, pruebas de hipótesis y análisis de variancia.

e) *Integración de lo estadístico y el contexto*.

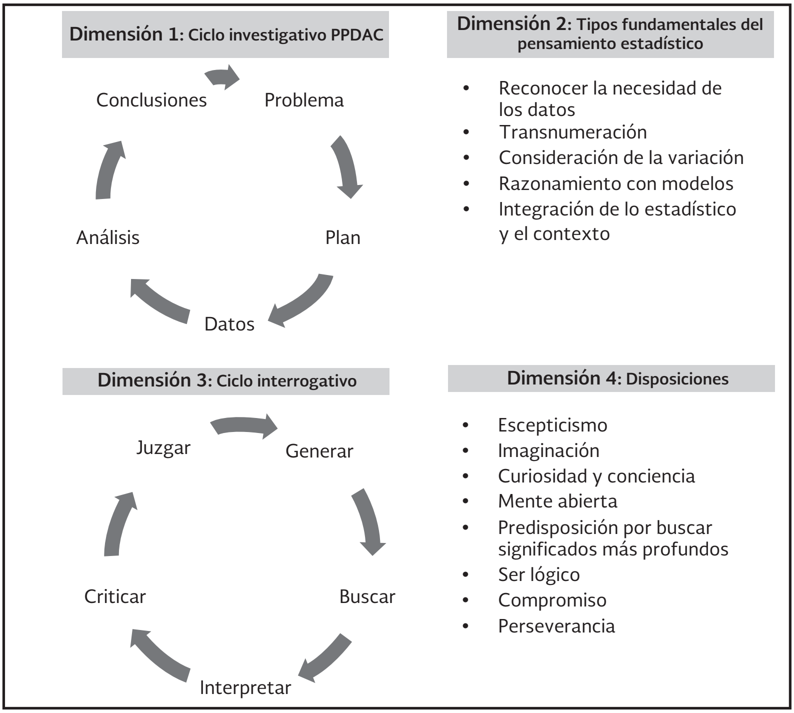


Figura 15. Dimensiones del modelo del pensamiento estadístico (Wild y Pfannkuch 1999).

El ciclo interrogativo se refiere a los procesos de producción y examen de ideas con relación al problema en una situación dada; éstos consisten en la generación de ideas, búsqueda de claves e indicios en los datos, interpretación de patrones o irregularidades, crítica de los diferentes aspectos del proceso, para finalmente, emitir juicios sobre la calidad y potencia del proceso y sus resultados. Finalmente, las disposiciones tienen que ver con las actitudes frente al trabajo estadístico y profesional. Estas dimensiones, así como sus contenidos, constituyen una información fundamental para el diseño de un currículo que contenga entre sus objetivos la formación de competencias estadísticas y para el profesor que asume la responsabilidad de promoverlas.

Las dimensiones del pensamiento estadístico están en un nivel superior al contenido estadístico, pues orientan acerca de cómo utilizarlo de una manera provechosa en la resolución de problemas.

Entre las características del modelo anterior y las unidades de contenido específico que suelen encontrarse en los currículos se localizan las ideas (estadísticas) fundamentales.

Las ideas estadísticas fundamentales

Un currículo basado en competencias debiera privilegiar las situaciones socialmente relevantes sobre contenidos específicos, esto implica renunciar a incluir en el currículo el catálogo tradicional de los contenidos. Díaz-Barriga (2012) opina: … hay que entender que no se ha logrado superar del todo al currículo enciclopedista, tan denostado en el discurso educativo de la innovación debido a que se encuentra cargado de contenidos factuales" (p. 31).

Sin embargo, llevar a cabo una reducción radical de contenidos a favor de un enfoque por competencias es una decisión difícil de tomar y de realizar en la práctica. ¿Cómo enseñar estadística si no se sigue la secuencia de la estadística descriptiva de comenzar con gráficas y seguir con medidas de tendencia central, medidas de variación y, finalmente, con regresión y correlación?

Si los contenidos del catálogo han pasado la prueba del tiempo y se han mantenido en la opinión favorable de los científicos, los elaboradores de currículo, los profesores y los padres de familia, ¿cómo elegir unos contenidos por sobre otros? O, ¿cómo explicar su posible ausencia en el currículo? Este es un problema difícil de resolver, al menos en el currículo de las matemáticas. Por ejemplo, omitir en éste algunos contenidos, como números primos o polinomios, bajo la justificación de que no son relevantes en la vida social de las personas, sería una afrenta para la comunidad de los matemáticos y los educadores matemáticos. En cambio, en la investigación en educación estadística se ha avanzado en proporcionar (ya sea de manera consciente o no) una salida a esta problemática, la cual ha consistido en definir ideas (estadísticas) fundamentales para organizar el currículo, éstas pueden ser desarrolladas apoyándose en contenidos específicos, pero sin imponer la necesidad de cubrirlos todos, ni en un orden definido. Los investigadores en educación estadística Burril y Bielher (2011) propusieron las ideas o nociones siguientes:

- *Datos.* Los ladrillos con los que se construyen conocimientos derivados del análisis estadístico son los datos. Los problemas en los que la estadística presta ayuda son aquellos en los cuales se pueden coleccionar y analizar datos.
- *Variación.* La variabilidad está presente en cualquier fenómeno

y la estadística identifica, describe, mide, explica y/o predice dicha variabilidad. A la descripción o medida de la variabilidad se le llama variación.

- *Distribución*. Una distribución es el instrumento matemático idóneo para describir un conjunto de datos y permitir a los estadísticos observar y analizar la variación presente en ellos. Una distribución teórica es la descripción de un patrón de variación de una variable aleatoria.
- *Representación*. Los diagramas y representaciones gráficas permiten descubrir y destacar diferentes ángulos de los datos y sus distribuciones gracias a que permiten su visualización. La potencia de las representaciones se expresa cuando son utilizadas en el análisis, y no sólo como vehículos de comunicación.
- *Asociación y correlación*. Establecer relaciones entre variables es un objetivo que suele presentarse en todas las disciplinas científicas y la estadística aporta los instrumentos para analizar los distintos grados de asociación o correlación que pueden presentarse entre variables estocásticas (categóricas o numéricas).
- *Probabilidad*. En la caracterización de Moore de la estadística, éste menciona tres elementos: datos, variación y azar; estos últimos se modelan a través de la probabilidad. En cualquier resultado de la estadística se presenta incertidumbre que se describe y mide mediante modelos probabilísticos.
- *Muestreo e inferencia*. La posibilidad de hacer afirmaciones con cierto grado de certeza acerca de una población con base en el análisis de una muestra es un objetivo fundamental de la estadística. Los instrumentos derivados de las ideas anteriores se articulan para llevar a cabo este propósito.

La presencia de estas ideas en el currículo no es suficiente para pensarlo como dirigido hacia el desarrollo de competencias, pues se puede constatar que en gran medida tales ideas ya han estado en el currículo tradicional; sin embargo, sí ofrecen una orientación para reducir la lista de contenidos usualmente presentes en el currículo. En síntesis, la estructura del currículo tradicional no es la apropiada para el desarrollo de competencias, pero tampoco un nuevo currículo podría ser un catálogo reducido que contuviera ahora sólo las ideas fundamentales. Sin embargo, si se tienen en cuenta y se agrega la propuesta de enfocar la enseñanza de la estadística hacia la realización de investigaciones y/o proyectos, entonces se podría ir configurando la forma de un nuevo currículo. Un intento serio en este sentido se encuentra en el marco curricular para la educación estadística de Franklin *et al.* (2005).

Un marco curricular para la educación estadística

Franklin *et al.* (2005) proponen un marco conceptual para la educación estadística de niveles preuniversitarios, el cual tiene por objetivo último propiciar el desarrollo de la competencia estadística: "nuestras vidas están gobernadas por números. Cualquier estudiante egresado del bachillerato debe ser capaz de usar el razonamiento estadístico para enfrentarse de manera inteligente con los requerimientos de la ciudadanía, el empleo y la familia, y estar preparado para una vida productiva saludable y feliz" (p. 1). Estos autores manifiestan que el marco que proponen es un complemento de las sugerencias curriculares que sobre el tema de análisis de datos y probabilidad presentan los Principios y Estándares de la NCTM (NCTM 2000). En su propuesta se enfatiza la diferencia entre matemáticas y estadística, y se señala a la variabilidad y al papel del contexto como factores que reflejan esa diferencia.

El marco mencionado está organizado en torno a dos nociones centrales: 1) la solución de problemas estadísticos; y 2) el papel de la variabilidad en los procesos de solución de problemas estadísticos.

Los procesos de solución de problemas estadísticos involucran cuatro componentes:

1. La formulación de una pregunta.
2. La recolección de datos.
3. El análisis de los datos.
4. La interpretación de los resultados.

Los autores sugieren que en cada uno de estos cuatro componentes debe intervenir la variabilidad. Por ejemplo, al hacer una pregunta conviene anticiparla. Mientras que en la recolección de datos se debe reconocer el papel crucial que representa, de tal manera que en algunos casos convendrá propiciarla, y en otros tratar de reducirla. En el análisis se intentará explicar la variabilidad en los datos; y, finalmente, en la interpretación se incluirá a la variabilidad como un elemento presente de diversas formas.

El marco conceptual de Franklin *et al.* (2005) orienta sobre cómo organizar los procesos de solución en las situaciones- problema que se consideren socialmente relevantes, y no pone en primer plano los contenidos que tradicionalmente han sido específicos de la disciplina. Su orientación indica el desarrollo de un pensamiento estadístico como objetivo transversal.

LA PROPUESTA DE ENSEÑANZA MEDIANTE PROYECTOS ESTADÍSTICOS

La importancia de investigaciones y proyectos estadísticos

Como ya se ha mencionado, la enseñanza de la estadística a través de proyectos (Díaz *et al*., 2008; MacGillivray y Pereira-Mendoza, 2011) se propone desarrollar el pensamiento estadístico, y la comunidad de educadores estadísticos han propuesto metodologías de enseñanza a tono con un currículo basado en el desarrollo de competencias. Esta propuesta presupone que la adquisición de contenidos específicos clave en la disciplina se realizará en el contexto de una investigación, lo que permitirá que el estudiante les asigne un sentido, se dé cuenta de su importancia y pueda utilizarlos en el desarrollo de otros proyectos. La enseñanza de la estadística a través de proyectos está basada en la instrumentación del marco curricular de Franklin *et al*. (2005). El carácter transversal de la propuesta consiste en que un proyecto de investigación estadística pueda realizarse de manera significativa y valiosa en el contexto de cualquier contenido disciplinar. Adicionalmente, esta metodología también es propicia para desarrollar otras competencias, como lo sugieren Díaz y colaboradores (2008).

Un proyecto para la enseñanza es una investigación en la que se trata de responder una pregunta mediante la recogida y análisis de datos dentro de las posibilidades que tienen los alumnos del grado correspondiente. Una condición para saber si una pregunta es adecuada es cuestionarse si genera un proceso investigativo:

1. La formulación de una pregunta.
2. La recolección de datos.
3. El análisis de los datos.
4. La interpretación de los resultados.

La pregunta debe contar con las siguientes condiciones:

a) La pregunta o problema deben ser de interés.
b) Los datos deben ser accesibles.
c) Los conceptos, técnicas y razonamientos que genera deber estar al nivel de los estudiantes.

La pregunta o problema deben ser de interés. Muchas preguntas se pueden responder con datos, pero pueden carecer de interés. Es cierto que un tema puede interesar a algunos estudiantes y a otros no, y pocos pueden ser los temas que a todos les interesen. Por eso conviene, al formular una pregunta, responder también ésta: ¿por qué es importante responderla? Motivando el interés de quienes en principio no les es atractivo. Otro aspecto que se debe cuidar es la concreción. Al respecto, Batanero y Díaz comentan:

> La fase de planteamiento de preguntas es una de las más difíciles. Los alumnos rara vez comienzan con un problema claramente formulado. Generalmente podrían comenzar sin preguntas claramente definidas y el papel del profesor es ayudarles a pasar de un tema general (deportes) a una pregunta que pueda contestarse (en la pasada temporada, ¿los equipos de fútbol que jugaron en su propio campo, lo hicieron mejor que los que jugaron en campo contrario? Batanero y Díaz, (2011).

Los datos deben ser accesibles. Los datos de un proyecto pueden ser recogidos por los propios estudiantes mediante una encuesta o la observación de algún fenómeno, o también pueden ser obtenidos de bases de datos. Hay muchas bases de datos de instituciones confiables como la Unesco. Por ejemplo, una tarea importante que debiera ser parte del curso de Estadística es conocer y explorar la página del INEGI (http://www.inegi.org. mx/) de manera que los estudiantes aprendan sobre los temas que hay en ella y los datos que ahí se pueden encontrar.

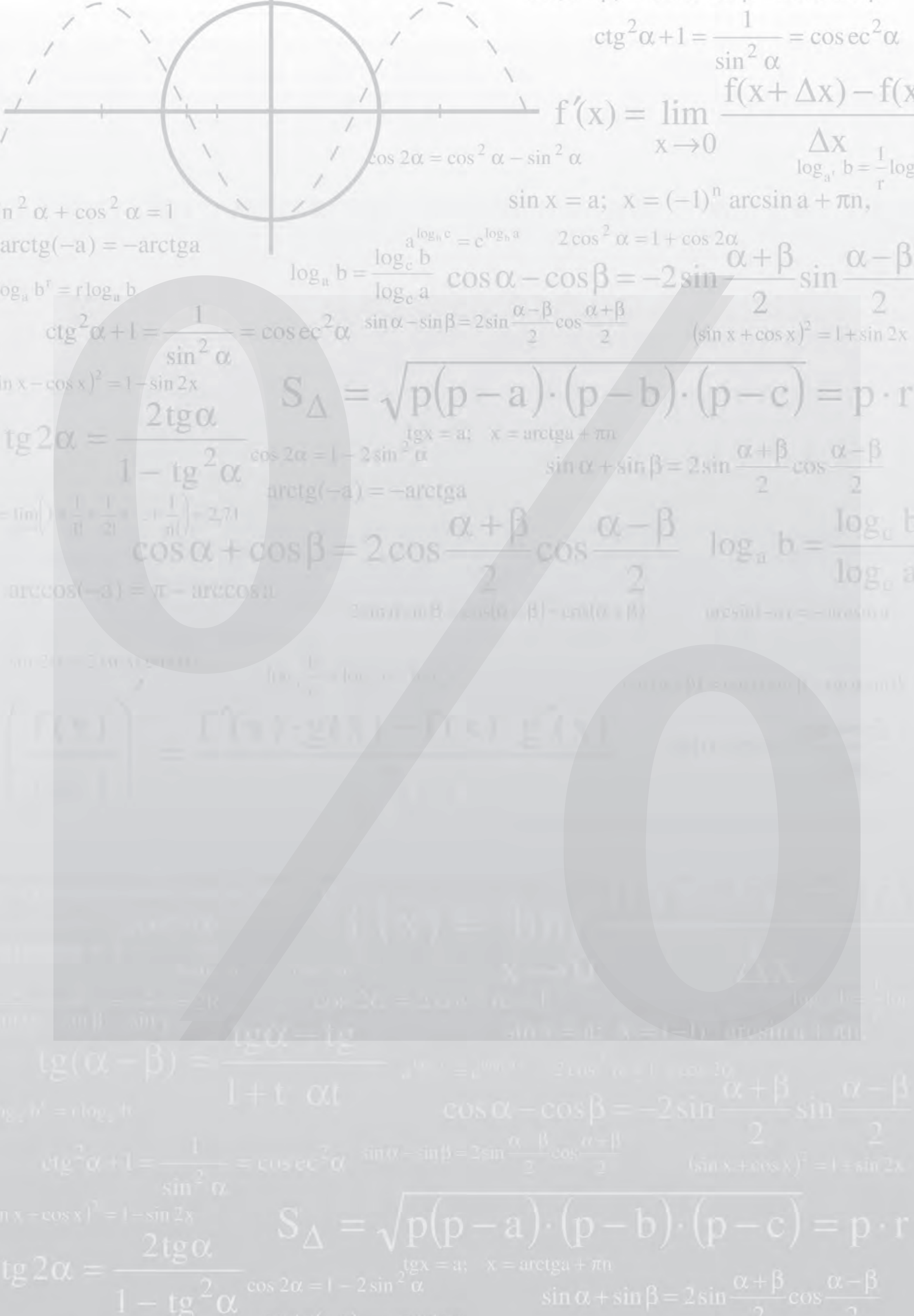
Los conceptos, técnicas y razonamientos deben estar al nivel de los estudiantes. El aspecto más difícil es el de prever que el tipo de conceptos, técnicas y razonamiento estadísticos del proyecto elegido esté al alcance del estudiante. Se recomienda la fórmula que se plantea en la propuesta de resolución de problemas: el problema no debe ser tan fácil que se resuelva mediante rutina, ni tan difícil que sea imposible resolverlo por el estudiante. Lo que hay que evitar es que el objetivo final sea aprender la técnica o el razonamiento, debe tenerse en cuenta que el objetivo debe ser decir algo acerca de la situación de la cual provienen los datos y para eso se requieren ciertos conceptos, técnicas y razonamientos.

TEMAS DE PROYECTOS

Conviene que el profesor tenga una lista de posibles proyectos a desarrollar por los estudiantes, haga un análisis de los conceptos, técnicas y razonamientos que implique, que pruebe la accesibilidad de los datos y la viabilidad para ser realizados en clase. Por ejemplo, los siguientes dan una idea de alguno de ellos; sin embargo, para las condiciones específicas de cada aula se pueden determinar muchos otros:

- ¿Existe discriminación laboral respecto a la mujer?
- ¿Qué posición (económica, demográfica, política o social) tiene México en América Latina?
- ¿Tiene ventaja el equipo de fútbol que juega en su propio campo?

En el siguiente capítulo se presenta el desarrollo completo de un proyecto que proponen Batanero, Díaz y Gea (2012), dicho proyecto ilustra una manera diferente de concebir la enseñanza de la estadística.



4. LAS ESTADÍSTICAS DE POBREZA Y DESIGUALDAD[1]

Objetivos

Se trata de analizar una serie de variables demográficas, entender su utilidad y el motivo por el cual son recogidas, analizar las relaciones entre las diferentes variables y estudiar las diferencias en sus distribuciones entre países, según su nivel de desarrollo.

Un objetivo importante es mostrar la utilidad de la estadística en el estudio de interrelaciones entre variables. Los alumnos, alternativamente, podrían tomar otros ficheros de datos de Internet o de anuarios estadísticos y desarrollar proyectos sobre otros temas en diferentes áreas de aplicación. El análisis es descriptivo, con finalidad exploratoria y no se plantean problemas de inferencia.

Rouncenfield (1995) presenta este proyecto sobre el cual los alumnos pueden trabajar en grupos, comparando las variables en los diferentes grupos de países y formulando por sí mismos preguntas de su interés. Mostraremos algunos ejemplos de las posibles actividades a desarrollar con los estudiantes.

Alumnos

Este proyecto podría ser desarrollado con alumnos de bachillerato o en un primer curso de estadística en la universidad. En este último caso, se podrían plantear preguntas de tipo inferencial, lo que requeriría el uso de procedimientos estadísticos más avanzados.

Los datos

La actividad se desarrolla en torno a un proyecto a partir de un fichero que contiene datos de 97 países y que ha sido adaptado del preparado por Rouncenfield (1995), quien usó como fuentes Day (1992) y Unesco (1990). Este fichero ha sido tomado de Journal of Statistical Education (www.amstat.org/publications/jse/). Contiene las siguientes variables, que se refieren a 1990:

- *Tasa de natalidad*: Niños nacidos vivos en el año por cada 1000 habitantes;
- T*asa de mortalidad*: Número de muertes en el año por cada 1000 habitantes;

[1] Carmen Batanero, Carmen Díaz y M. Magdalena Gea

- *Mortalidad infantil*: Número de muertes en el año por cada 1000 niños de menos de 1 año;
- *Esperanza de vida* al nacer para hombres y mujeres;
- *PNB*: Producto Nacional Bruto per cápita en dólares (EU);
- *Grupo*: Clasificación de países en función de la zona geográfica y situación económica, en las siguientes categorías: 1 = Europa Oriental, 2 = Iberoamérica, 3 = Europa Occidental, América del Norte, Japón, Australia, Nueva Zelanda, 4 = Oriente Medio, 5 = Asia, 6 = África.

Hemos añadido el número de habitantes en 1990 en miles de personas (población), tomado del anuario publicado por el periódico español *El País*. En la Tabla 2 listamos los datos del proyecto.

Tabla 2. Fichero de datos del proyecto "Análisis demográfico". (*continúa*)

País	Grupo	Tasa Natalidad	Tasa Mortalidad	Mortalidad Infantil	Esperanza de vida hombre	Esperanza de vida mujer	PNB	Población (miles)
Afganistán	5	40.4	18.7	181.6	41.0	42.0	168	16000
Albania	1	24.7	5.7	30.8	69.6	75.5	600	3204
Alemania Oeste	3	11.4	11.2	7.4	71.8	78.4	22320	16691
Alemania Este	1	12.0	12.4	7.6	69.8	75.9		61337
Argelia	6	35.5	8.3	74.0	61.6	63.3	2060	24453
Angola	6	47.2	20.2	137	42.9	46.1	610	9694
Arabia Saudí	4	42.1	7.6	71.0	61.7	65.2	7050	13562
Argentina	2	20.7	8.4	25.7	65.5	72.7	2370	31883
Austria	3	14.9	7.4	8.0	73.3	79.6	17000	7598
Bahréin	4	28.4	3.8	16.0	66.8	69.4	6340	459
Bangladesh	5	42.2	15.5	119.0	56.9	56.0	210	111590
Bélgica	3	12.0	10.6	7.9	70.0	76.8	15540	9886
Bielorusia	1	15.2	9	13.1	66.4	75.9	1880	
Bolivia	2	46.6	18.0	111.0	51.0	55.4	630	7110
Botswana	6	48.5	11.6	67.0	52.3	59.7	2040	1217
Brasil	2	28.6	7.9	63.0	62.3	67.6	2680	147294
Bulgaria	1	12.5	11.9	14.4	68.3	74.7	2250	9001
Camboya	5	41.4	16.6	130.0	47.0	49.9		8250
Canadá	3	14.5	7.3	7.2	73.0	79.8	20470	26302
Colombia	2	27.4	6.1	40.0	63.4	69.2	1260	32335
Congo	6	46.1	14.6	73.0	50.1	55.3	1010	2208
Corea (Norte)	5	23.5	18.1	25.0	66.2	72.7	400	21143
Checoslovaquia	1	13.4	11.7	11.3	71.8	77.7	2980	15641
Chile	2	23.4	5.8	17.1	68.1	75.1	1940	12980
China	5	21.2	6.7	32.0	68.0	70.9	380	1105067
Dinamarca	3	12.4	11.9	7.5	71.8	77.7	22080	5132
Ecuador	2	32.9	7.4	63.0	63.4	67.6	980	10329
Egipto	6	38.8	9.5	49.4	57.8	60.3	600	51390
Emiratos Árabes	4	22.8	3.8	26.0	68.6	72.9	19860	1544
España	3	10.7	8.2	8.1	72.5	78.6	11020	39161
Etiopía	6	48.6	20.7	137	42.4	45.6	120	48861
Filipinas	5	33.2	7.7	45.0	62.5	66.1	730	61224
Finlandia	3	13.2	10.1	5.8	70.7	78.7	26040	4974
Francia	3	13.6	9.4	7.4	72.3	80.5	19490	56119
Gabón	6	39.4	16.8	103.0	49.9	53.2	390	1105
Gambia	6	47.4	21.4	143.0	41.4	44.6	260	848
Ghana	6	44.4	13.1	90.0	52.2	55.8	390	14425
Grecia	3	10.1	9.2	11.0	65.4	74.0	5990	10039
Guayana	2	28.3	7.3	56.0	60.4	66.1	330	95
Holanda	3	13.2	8.6	7.10	73.3	79.9	17320	14828
Hong Kong	5	11.7	4.9	6.10	74.3	80.1	14210	5735
Hungría	1	11.6	13.4	14.8	65.4	73.8	2780	10587
India	5	30.5	10.2	91.0	52.5	52.1	350	832535
Indonesia	5	28.6	9.4	75.0	58.5	62.0	570	178211
Irán	4	42.5	11.5	108.1	55.8	55.0	2490	50204

Tabla 2. Fichero de datos del proyecto "Análisis demográfico". (*continuación*)

País	Grupo	Tasa Natalidad	Tasa Mortalidad	Mortalidad Infantil	Esperanza de vida hombre	Esperanza de vida mujer	PNB	Población (miles)
Iraq	4	42.6	7.8	69.0	63.0	64.8	3020	18271
Irlanda	3	15.1	9.1	7.5	71.0	76.7	9550	3537
Israel	4	22.3	6.3	9.7	73.9	77.4	10920	4525
Italia	3	9.7	9.1	8.8	72.0	78.6	16830	57537
Japón	3	9.9	6.7	4.0	75.9	81.8	25430	123045
Jordania	4	38.9	6.4	44.0	64.2	67.8	1240	4041
Kenia	6	47.0	11.3	72.0	56.5	60.5	370	23277
Kuwait	4	26.8	2	15.6	71.2	75.4	16150	2020
Líbano	4	31.7	8.7	48.0	63.1	67.0		2900
Libia	6	44.0	9.4	82.0	59.1	62.5	5310	4395
Malasia	5	31.6	5.6	24.0	67.5	71.6	2320	17340
Malawi	6	48.3	25.0	130.0	38.1	41.2	200	8230
Marruecos	6	35.5	9.8	82.0	59.1	62.5	960	24567
México	2	29.0	23.2	43.0	62.1	66.0	2490	85440
Mongolia	5	36.1	8.8	68.0	60.0	62.5	110	2128
Mozambique	6	45.0	18.5	141	44.9	48.1	80	15357
Namibia	6	44.0	12.1	135.0	55.0	57.5	1030	1300
Nepal	5	39.6	14.8	128.0	50.9	48.1	170	18431
Nigeria	6	48.5	15.6	105.0	48.8	52.2	360	113665
Noruega	3	14.3	10.7	7.8	67.2	75.7	23120	4215
Omán	4	45.6	7.8	40.0	62.2	65.8	5220	1486
Pakistán	5	30.3	8.1	107.7	59.0	59.2	380	109950
Paraguay	2	34.8	6.6	42.0	64.4	68.5	1110	4161
Perú	2	32.9	8.3	109.9	56.8	66.5	1160	21142
Polonia	1	14.3	10.2	16.0	67.2	75.7	1690	38061
Portugal	3	11.9	9.5	13.1	66.5	72.4	7600	10333
Rumania	1	13.6	10.7	26.9	66.5	72.4	1640	23148
Sierra Leona	6	48.2	23.4	154.0	39.4	42.6	240	4040
Singapur	5	17.8	5.2	7.5	68.7	74.0	11160	2664
Somalia	6	50.1	20.2	132.0	43.4	46.6	120	6089
Sri Lanka	5	21.3	6.2	19.4	67.8	71.7	470	16779
Sudáfrica	6	32.1	9.9	72.0	57.5	63.5	2530	34925
Sudán	6	44.6	15.8	108.0	48.6	51.0	480	24423
Suecia	3	14.5	11.1	5.6	74.2	80.0	23660	8485
Suiza	3	12.5	9.5	7.1	73.9	80.0	34064	6541
Suazilandia	6	46.8	12.5	118.0	42.9	49.5	810	761
Tailandia	5	22.3	7.7	28.0	63.8	68.9	1420	55200
Tanzania	6	50.5	14.0	106	51.3	54.7	110	25627
Túnez	6	31.1	7.3	52.0	64.9	66.4	1440	7988
Turquía	4	29.2	8.4	76.0	62.5	65.8	1630	54899
U.K.	3	13.6	11.5	8.4	72.2	77.9	16100	57270
EU	3	16.7	8.1	9.1	71.5	78.3	21790	248243
Ucrania	1	13.4	11.6	13.0	66.4	74.8	1320	
Uganda	6	52.2	15.6	103.0	49.9	52.7	220	16722
Uruguay	2	18.0	9.6	21.9	68.4	74.9	2560	3067
URSS	1	17.7	10.0	23.0	64.6	74.0	2242	287664
Venezuela	2	27.5	4.4	23.3	66.7	72.8	2560	19244
Vietnam	5	31.8	9.5	64.0	63.7	67.9		65758
Yugoslavia	1	14.0	9.0	20.2	68.6	74.5		23707
Zaire	6	45.6	14.2	83.0	50.3	53.7	220	34442
Zambia	6	51.1	13.7	80.0	50.4	52.5	420	7837
Zimbabue	6	41.7	10.3	66.0	56.5	60.1	640	9567

Preguntas, actividades y gestión de la clase

La actividad inicial consiste en discutir el significado de las variables de este fichero y analizar cómo se han calculado las diferentes tasas: natalidad, mortalidad, esperanza de vida, PNB. Los alumnos podrían investigar qué otros indicadores alternativos se emplean para obtener un indicador demográfico o económico de la riqueza de un país. El profesor podría pedirles que busquen artículos en la prensa en que se hable de alguno de estos indicadores y que expliquen con sus propias palabras la utilidad que pueden tener y que averigüen quién y cómo los calcula.

1. ¿Podrías explicar qué significa cada una de las variables del fichero? ¿Quién y cómo las calcula? ¿Cómo se recogen los datos? ¿Habría otro modo de calcular el índice de natalidad? ¿Para qué sirve? ¿Podrías encontrar alguna noticia en la prensa relacionada con estas variables?

Un punto muy importante es la discusión de las variables y el problema de la medición. Es preciso concienciar a los alumnos de la dificultad que reviste el proceso de categorización o de medición, porque la realidad es siempre más compleja que nuestros métodos para estudiarla. La toma de conciencia sobre la complejidad del proceso de elaboración de las estadísticas demográficas o económicas es un paso importante para valorar el trabajo del estadístico y fomentar la cooperación en censos y encuestas.

En este fichero se ha usado un código para agrupar los países en función de la zona geográfica y desarrollo económico. Los alumnos podrían sugerir otras variables de clasificación de los países o añadir otras variables o países al fichero. El trabajo con un fichero completo, en lugar de centrarse en variables aisladas, supone el inicio de una filosofía multivariante donde cada variable cobra su importancia, o bien, es explicada en función del resto y donde el alumno puede tratar de comprobar sus conjeturas con la incorporación de nuevas variables al estudio.

La figura 16 muestra un diagrama de barras donde representamos el número de países en los diferentes grupos de países. A partir del mismo es fácil construir una tabla de frecuencias y discutir el significado de las frecuencias absolutas, relativas y porcentajes. Los alumnos pueden analizar las ventajas que el diagrama de barras tiene frente a la tabla para visualizar el grupo que tiene mayor / menor número de países. Asimismo pueden elaborar otros gráficos adecuados para representar esta variable.

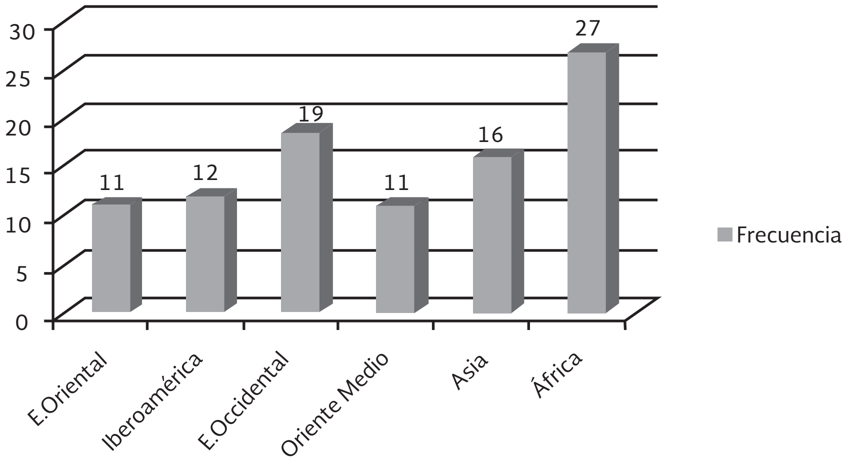


Figura 16. Número de países por grupo.

2. En las figuras 17 y 18 hemos representado el número promedio de habitantes en cada país, según grupo, usando dos promedios diferentes: media y mediana. ¿Por qué las dos gráficas son tan diferentes? ¿Elegirías la media o la mediana para representar el número típico de habitantes en los países, según la zona geográfica?

La elaboración de una tabla o un gráfico ya supone una primera reducción de los datos, pero a veces queremos hallar un único valor representativo de la distribución. Esta actividad puede realizarse a partir de las gráficas ya elaboradas o también pedir a los alumnos que las construyan previamente.

En el segundo caso, la clase puede dividirse en grupos para calcular estos promedios, así como la moda, y para explicar lo que representa cada uno de estos promedios y elegir en cada grupo el que mejor lo representa, argumentando la elección. Se puede pedir a los alumnos que señalen las principales diferencias entre los dos gráficos y que decidan cuál de los dos promedios acentúa más las diferencias explicando la razón

Mientras la media puede, en este ejemplo, interpretarse como el reparto equitativo (número de habitantes) entre todos los habitantes de cada zona geográfica, la mediana indica simplemente el número de habitantes que tuviesen tantos países con menos o más habitantes que él. La diferencia más importante se observa en Asia, porque como China e India tienen una población tan grande, al calcular la media, los valores de la variable en estos dos países hacen subir mucho el resultado. Los estudiantes pueden comprobarlo, haciendo el cálculo de media y mediana en los países asiáticos.

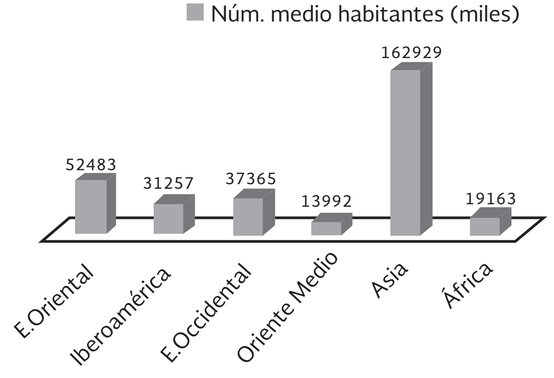


Figura 17. Media del número de habitantes en diferente zonas.

La actividad servirá para analizar algunas propiedades de la media aritmética y mediana. Mientras que la media es muy sensible a los valores extremos, pues todos los valores intervienen en el cálculo, la mediana es "robusta", pues en su cálculo no intervienen los valores, sino sólo el orden relativo de los mismos.

El profesor puede recordar el significado de un valor atípico. China e India tendrían valores atípicos de la población en este ejemplo. Finalmente, será importante hacer ver a los estudiantes la importancia de decidir adecuadamente la medida de tendencia central para representar los datos.

¡La media no sería en este caso un buen representante, pues pocos países asiáticos se verían representados si se dice que el número promedio de habitantes es 162 millones!

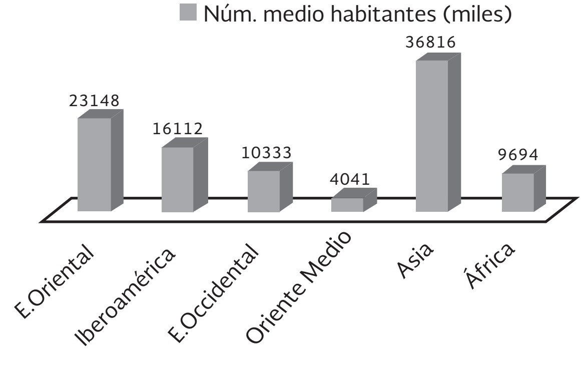


Figura 18. Mediana del número de habitantes en diferente zonas.

3. ¿Qué representa el valor obtenido al calcular la media de la esperanza media de vida al nacer en estos 97 países? ¿Cómo habría que hacer para calcular la esperanza media de vida al nacer en hombres y mujeres, si no tenemos en cuenta el país de nacimiento?

En este fichero las unidades estadísticas son agregados, lo que tiene repercusión en la interpretación de los promedios de las variables; por ejemplo, la media (simple) de todas las esperanzas de vida al nacer en los distintos países no es igual a la esperanza global de vida, sino que ésta tiene que ser calculada como una media ponderada en función del número de hombres/mujeres de cada país.

Esto puede servir de reflexión sobre las diferentes unidades estadísticas que se usan en un estudio, la necesidad que a veces tenemos de trabajar con valores aproximados y la forma de combinar estudios parciales para obtener índices globales, así como para reflexionar sobre las propiedades de los promedios y la comprensión de las mismas por parte de los estudiantes (Tormo, 1995; Batanero, 2000). Esta misma observación debe hacerse en la interpretación de otros estadísticos, como los de dispersión o los coeficientes de correlación.

El profesor podría animar a los estudiantes a consultar otros datos estadísticos sobre agregados producidos por el Instituto Nacional de Estadística y ver la dificultad que a veces tiene decidir los agregados. En nuestro ejemplo, Europa se ha dividido en dos partes; Oriental y Occidental, y algunos países como Estados Unidos o Japón se han añadido a Europa Occidental. Con ejemplos sencillos se podría conducir a los alumnos a la idea de media ponderada y hacerles ver la necesidad de tener en cuenta el número de hombres y mujeres de cada país para calcular el promedio.

4. Nosotros hemos calculado la esperanza media de vida global en hombres y mujeres, ponderando los datos de cada país por su número de habitantes (suponiendo un número aproximadamente igual de hombres y mujeres). En la figura 19 hemos representado la esperanza media de vida en hombres y mujeres con dos escalas diferentes. Compara estos dos gráficos e indica si te parecen o no adecuados para representar la diferencia entre la esperanza de vida media de mujeres y hombres. Uno de los dos gráficos obtuvo directamente del ordenador, mientras que el otro ha sido manipulado. Averigua cuál ha sido manipulado.

La finalidad de esta pregunta es concienciar a los estudiantes sobre la posibilidad de que una gráfica sea manipulada para distorsionar la información y también de la importancia de las escalas adecuadas en una gráfica. Una persona culta debiera poder leer críticamente los gráficos estadísticos que encuentra en la prensa, Internet, medios de comunicación y trabajo profesional. Esto supone no sólo la lectura literal del gráfico, sino identificar las tendencias y variabilidad de los datos, así como detectar los posibles errores conscientes o inconscientes que puedan distorsionar la información representada (Schield, 2006). Asimismo, debiera conocer los convenios de construcción de los diferentes tipos de gráficos y ser capaz de construir uno correctamente.

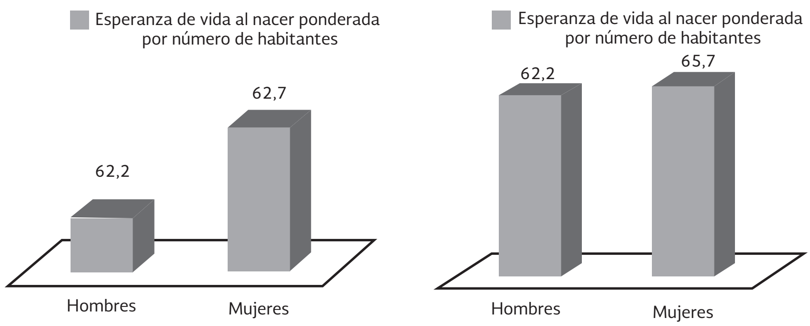


Figura 19. Esperanza de vida media de hombres y mujeres.

Monteiro y Ainley (2007) indican que la lectura de gráficos en el contexto escolar es una tarea más limitada que la posible interpretación de éstos en otras actividades de la vida diaria. La razón dada por los autores es que, mientras en la escuela sólo pedimos a los estudiantes una respuesta correcta desde el punto de vista matemático, en contextos extraescolares intervienen también otros conocimientos no matemáticos. En esta actividad tratamos de que los estudiantes sean críticos en la lectura de gráficos similares a los que puedan encontrar en la vida diaria y también con el uso de las opciones por defecto del *software* estadístico, que no son siempre las más adecuadas.

5. Construye ahora una tabla de frecuencias que muestre la distribución de las tasas de natalidad.

   ¿Por qué en este caso conviene agrupar en intervalos?

   ¿Qué representa la frecuencia dentro de un intervalo?

   ¿Cuántos intervalos conviene usar en la tabla de frecuencias? ¿Cómo representarías gráficamente estos datos? ¿Es simétrica la distribución? ¿Cómo cambia la representación gráfica al variar el número de intervalos? ¿Y si usamos intervalos de distinta amplitud? ¿Qué significa y cómo representarías la frecuencia acumulada? ¿Qué posición ocupa tu país respecto a la tasa de natalidad? ¿Hay algunos países atípicos respecto a la tasa de natalidad? Para contestar esta y otras preguntas similares puedes usar el gráfico del tallo y hojas (figura 20).

Cuando la variable presenta un número grande de valores, surge la necesidad de agrupación. Ahora bien, no existe una regla fija sobre la forma de construir los intervalos de clase, aunque se recomienda un número de intervalos aproximadamente igual a la raíz cuadrada del número de datos. En esta actividad se muestra un ejemplo de cómo en estadística es posible tener más de una solución correcta al mismo problema y también cómo la forma del histograma varía en función de la elección de los intervalos.

Junto con las representaciones gráficas tradicionales, como diagramas de barras o histogramas, en el análisis exploratorio de datos aparecen nuevas representaciones como el diagrama de tallo y hojas o los gráficos de cajas, cuya potencia exploratoria se acentúa con el paso de una a otra representación, así como la selección de partes del fichero para realizar estudios comparativos; por ejemplo, al comparar las variables en los distintos grupos de países, que es el tema de la siguiente actividad.

El interés no sólo se centra en las tendencias, sino en la variabilidad, así como el estudio de los valores atípicos. Vemos también cómo la idea de distribución siempre es relativa a un colectivo; por eso un valor puede ser atípico dentro de un subconjunto de datos y no serlo en el global.

```
0|99
1|00111122222333333444444
1|556778
2|011222334
2|677888899
3|00111122234
3|5568899
4|011122224444
4|5556667778888 8
5|0012
```

Figura 20. Gráfico de tallo y hojas: tasa de natalidad.

Este diagrama conserva los valores originales de los datos y a la vez muestra la frecuencia de valores, así como la moda. Es muy sencillo de construir y de determinar, a partir de él, la mediana y los cuartiles. Puede ser un paso previo a la construcción de una tabla de frecuencias o un histograma o bien al diagrama de frecuencias acumuladas (figura 21). El diagrama de frecuencias acumuladas servirá para determinar gráficamente la mediana y para introducir la idea de percentiles de la distribución. Será interesante para cada alumno analizar en qué percentil se encuentra su país de origen respecto a cada una de las variables.

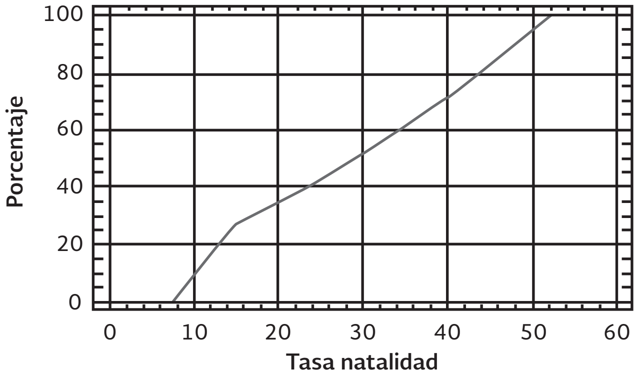


Figura 21. Gráfico de frecuencias acumuladas.

6. Diferencias en la tasa de natalidad. Una vez estudiada globalmente la tasa de natalidad dividiremos la clase en grupos para estudiar si es la misma en los diferentes grupos de países. Investigaremos qué grupos de países tienen valores atípicos para la tasa de natalidad y qué países tienen una tasa de natalidad atípica respecto a su grupo, así como por qué al realizar un gráfico global en todo el fichero no aparecen valores atípicos y sí se muestran dentro de los grupos. Se pide listar todas las diferencias observadas en los diversos grupos de países.

Países del grupo 1

1|1
1|22333
1|445
1|7

HI|24.7

Países del grupo 2

1|8
2|03
2|77889
3|224

HI|46.6

Países del grupo 3

0|99
1|1111
1|222
1|3333
1|4444
1|5
1|6

Países del grupo 4

2|22
2|689
3|1
3|8
4|222
4|5

Países del grupo 5

1|1
1|7
2|1123
2|8
3|00113
3|69
4|012
3|89

Países del grupo 6

LO|31.1

3|
3|
3|2
3|55
3|
4|1
4|444455
4|66777

Figura 22. Tasa de natalidad en los diferentes grupos de paises.

Los alumnos elaborarán gráficos como los presentados en las figuras 22, 23, 24. El gráfico de tallo y hojas ha sido útil para analizar las diferencias en la tasa de natalidad, pero también podemos usar otras representaciones, como diagramas de puntos y gráficos de caja. Divididos en grupos, los alumnos pueden tratar de hallar representaciones alternativas que pongan de manifiesto las diferencias en la tasa de natalidad. Por ejemplo, podemos observar las figuras 23 y 24 y comparar con los gráficos del tronco, señalando las ventajas que tiene cada una de las representaciones gráficas y cómo podríamos cambiar cada gráfica para resaltar más (o menos) las diferencias.

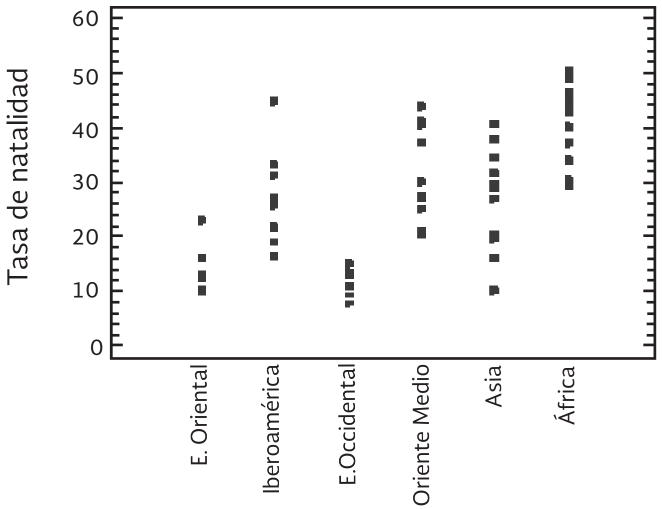


Figura 23. Gráfico de puntos.

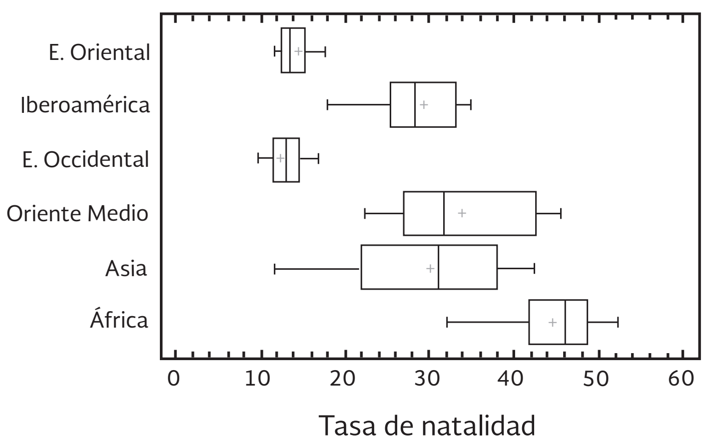


Figura 24. Gráfico de caja.

7. Esperanza de vida en hombres y mujeres. Se dice, por ejemplo, que las mujeres forman el sexo débil. Sin embargo, muchas chicas estarían en contra de esta opinión cuando comparen en algunos países la esperanza de vida al nacer en hombres y mujeres.

   ¿Hay mayor variabilidad entre países en la esperanza de vida en hombres o en mujeres? ¿En qué porcentaje de países la esperanza de vida de hombres (mujeres) es mayor de 58 años?

¿y de 68 años? ¿Cuál es el valor de la esperanza de vida de modo que 70% de países tiene una esperanza de vida mayor? ¿Es igual en hombres y mujeres? ¿Cómo podrías disimular la diferencia de esperanzas de vida en hombres y mujeres en cada gráfico?

En la actividad anterior hemos comparado una variable en distintos subconjuntos de países, pero a veces tiene también sentido comparar dos variables diferentes en el total de los datos. Podríamos investigar si la distribución de la esperanza de vida en el total de países es igual en hombres y mujeres analizando para ello los gráficos de las figuras 25 a 26 y contestar las preguntas anteriores.

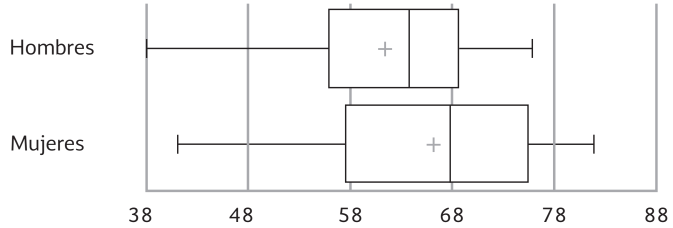


Figura 25. Gráficos de caja. Esperanza de vida en hombres y mujeres.

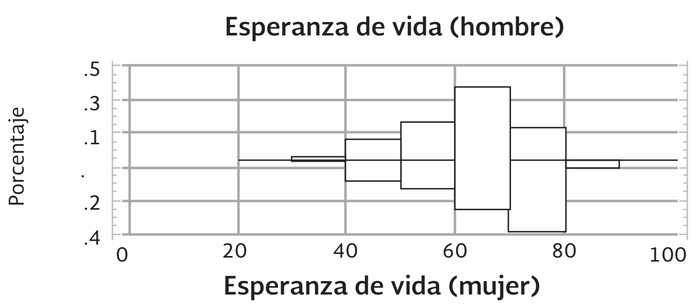


Figura 26. Histogramas. Esperanza de vida en hombres y mujeres.

En esta actividad se vuelven a aplicar los conceptos aprendidos al estudio de las diferencias de dos variables en una misma muestra. Se aplica también el estudio de las frecuencias acumuladas, percentiles y sus rangos. Hacemos notar que el cálculo de percentiles y sus rangos puede hacerse directamente a partir de la gráfica sin tener que recurrir a las fórmulas de cálculo que son complejas y para las

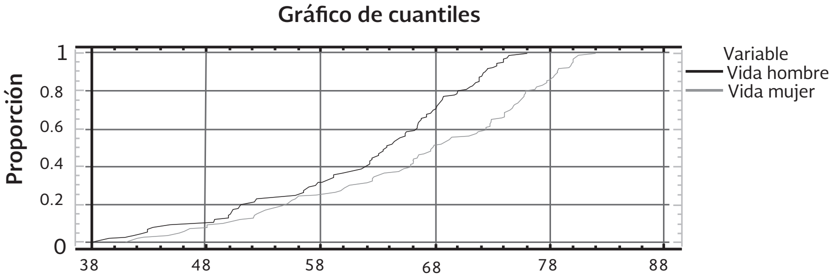


Figura 27. Distribución acumulativa de la esperanza de vida en hombres y mujeres.

que existe una variedad de casos, según las variables estén o no agrupadas en intervalos.

Actividades de ampliación

8. Relación de la esperanza de vida con otras variables. En la figura 28 hemos representado los diagramas de dispersión de la esperanza de vida del hombre en cada país, en función de diversas variables. ¿Qué variables están relacionada con la esperanza de vida del hombre? ¿En cuál es la relación directa o inversa?

   ¿Cuáles influyen o son influidas por la esperanza de vida del hombre? ¿Cuál sirve mejor para predecir la esperanza de vida? ¿En qué casos la relación podría ser debida a otras variables? ¿Podríamos en alguno de los casos hallar una función matemática para predecir, aproximadamente, la esperanza de vida del hombre a partir de la otra variable? ¿Qué tipo de función?

Al analizar las relaciones entre dos variables numéricas los alumnos deben de extender la idea de dependencia funcional a dependencia aleatoria y diferenciar sus tipos (lineal o no; directa e inversa), así como graduar, al menos intuitivamente, la intensidad de la relación. Es importante también diferenciar correlación y causalidad y analizar los distintos tipos de relaciones que pueden llevar a la existencia de correlación: dependencia causal; interdependencia, dependencia indirecta, concordancia y correlación espuria (Estepa, 1995). Una vez detectada la correlación, el interés se centra en la búsqueda de modelos que puedan predecir las variables explicadas en función de las variables explicativas, lo que de nuevo conecta con otro contenido del currículo de matemáticas: las funciones.

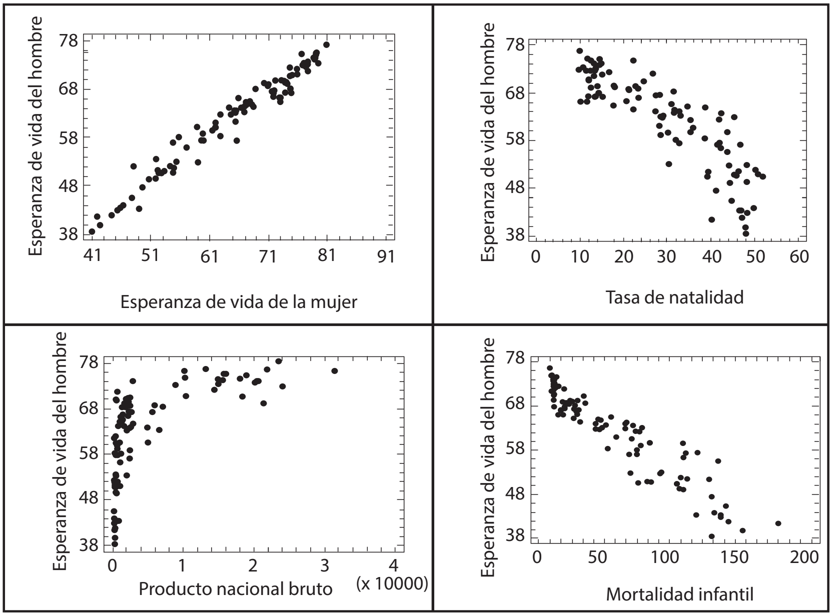


Figura 28. Relación de la esperanza de vida del hombre con otras variables.

9. Regresión. Estudiemos con más detalle la tasa de natalidad en función del PNB. ¿Cuál es el valor del coeficiente de correlación? ¿Qué te indica? Si utilizamos una recta para predecir la tasa de natalidad en función del PNB, ¿sería un buen modelo de predicción? ¿Podríamos mejorar la predicción cambiando el tipo de función?

Con alumnos universitarios se puede proceder al estudio más formal de la regresión, primero calculando el coeficiente de correlación, que en este caso (Tabla 3) es alto y negativo.

No obstante el alto valor de coeficiente, por sí solo no es indicativo de una relación lineal entre las variables, por lo que conviene representar gráficamente los datos en un diagrama de dispersión (figura 29). En el diagrama de dispersión y el coeficiente de correlación se observa una relación inversa, ya que a mayor tasa de natalidad menor PNB, pero no lineal.

Tabla 3. Resumen del modelo.

Modelo	R	R cuadro	R. cuadro corregida	Error típ. de la estimación
1	-.629(a)	.396	.389	10.70878

(a) Variables predictoras: (Constante), Producto nacional bruto en millones de dólares.

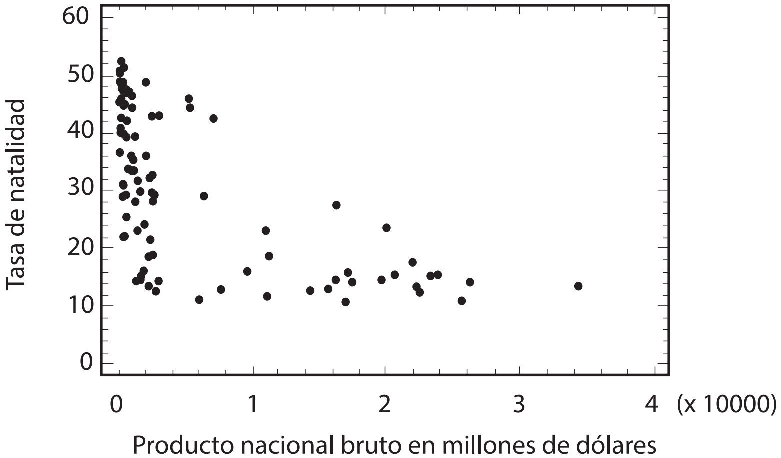


Figura 29. Diagrama de dispersión.

Tabla 4. Resumen del modelo y estimaciones de los parámetros.

	Resumen del modelo					Estimaciones de los parámetros			
Ecuación	R^2	F	gl1	gl2	Sig.	Constante	b1	b2	b3
Lineal	.396	58.282	1	89	.000	35.537	-0.001		
Logarítmica	.542	105.278	1	89	.000	75.521	-6.132		
Cuadrática	.454	36.645	2	88	.000	37.737	-.002	5.49E-8	
Cúbica	.481	26.898	3	87	.000	39.644	-.004	2.19E-3	-3.62E-12
Potencia	.553	110.165	1	89	.000	157.069	-.240		
Exponencial	.446	71.554	1	89	.000	33.352	-4.37E-05		

a La variable independiente es producto nacional bruto en millones de dólares

La proporción de la varianza de la tasa de natalidad explicada por el PNB con un modelo lineal es 38.90 %. Luego se puede tratar de comparar otros posibles modelos, es decir, otras posibles curvas que se ajusten mejor a los datos que un modelo lineal (recta). Estudiando otros modelos diferentes (tabla 4), el que mayor porcentaje de varianza ofrece es el de potencia o el logarítmico. En el cuadro anterior obtenemos las estimaciones de los parámetros. Gráficamente vemos que el ajuste es mejor que el caso lineal (figura 30).

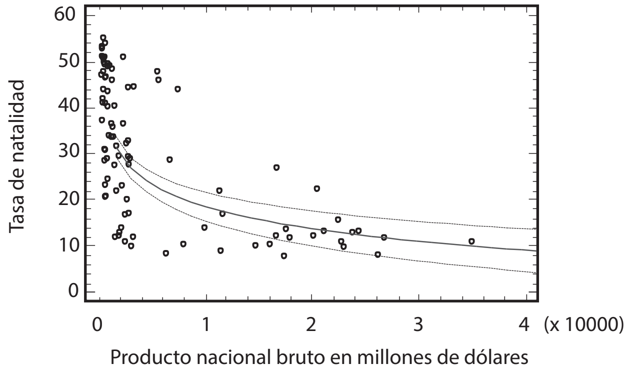


Figura 30. Ajuste de un modelo.

10. Estudiemos ahora qué otras variables se relacionan linealmente con la tasa de natalidad y tratemos de encontrar una función que prediga la tasa de natalidad en función de estas variables.

Estudiando las correlaciones (tabla 5) vemos que todas son significativas, siendo la más alta las esperanzas de vida (inversa) y a continuación la mortalidad infantil. La natalidad aumenta si hay mucha mortalidad infantil y poca esperanza de vida. Se puede intentar ajustar una ecuación con más de una variable dependiente, introduciendo paso a paso aquellas que tienen más correlación con la dependiente.

En la regresión por pasos (tabla 6) sólo se incluyen dos variables en la ecuación, que ya explican 83% de la varianza (esperanza de vida de la mujer y tasa de mortalidad).

El resto de las variables se excluye porque no hacen mejorar el modelo. Los coeficientes permiten escribir la ecuación de regresión que aparece en la (tabla 7). Prediciendo la esperanza de vida del hombre en función de la esperanza de vida de la mujer y la tasa de natalidad se obtiene un modelo suficientemente explicativo.

Tabla 5. Resumen del modelo y estimaciones de los parámetros.

		Tasa de natalidad
Tasa de mortalidad	Correlación de Pearson Sig.(bilateral)	,486(**) ,000
Mortalidad infantil	Correlación de Pearson Sig.(bilateral)	,858(**)
Esperanza de vida al nacer en el hombre	Correlación de Pearson Sig.(bilateral)	,867(**) ,000
Esperanza de vida al nacer en la mujer	Correlación de Pearson Sig.(bilateral)	,894(**) ,000
Producto nacional bruto en millones de dólares	Correlación de Pearson Sig.(bilateral)	,629(**) ,000

**La correlación es significativa al nivel 0,01 (bilateral).

Tabla 6. Resumen del modelo.

					Estadísticos de cambio				
Modelo	R	R^2	R^2 corregida	Error típ. estimación	Cambio R^2	Cambio en F	gl 1	gl 2	Sig. del cambio F
1	,894(a)	,800	,798	6,16040	,800	356,051	1	89	,000
2	,915(b)	,837	,833	5,60090	,037	19,669	1	88	,000

(a) Variables predictorias: (Constante), esperanza de vida al nacer en la mujer.
(b) Variables predictorias: (Constante), esperanza de vida al nacer en la mujer, tasa de mortalidad.

Tabla 7. Coeficientes (a).

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
2	(Constante)	126,602	6,558		19,304	,000
	Esperanza de vida al nacer en la mujer	-1,341	,076	-1,090	17,685	,000
	Tasa de mortalidad	-,799	,180	-,273	-4,435	,000

a Variable dependiente: tasa de natalidad.

Se podría completar el estudio con contrastes de diferencias de medias. La diferencia de esperanzas de vida en hombres y mujeres se podría estudiar mediante la prueba t de muestras relacionadas (tablas 8 y 9). La diferencia de cualquiera de las variables en las diferentes zonas también se podría estudiar mediante análisis de varianza de una vía (tablas 10 y 11).

Tabla 8. Estadísticos de muestras relacionadas.

		Media	N	Desviación típica	Error típ. media
Par 1	Esperanza de vida hombre	61,4856	97	9,61597	,97635
	Esperanza de vida mujer	66,1511	97	11,00539	1,11743

Tabla 9. Prueba de muestras relacionadas

Diferencias relacionadas					t	gl	Sig. (bilateral)
Media	Desviación típ.	Error típ. de la media	95% intervalo de confianza para la diferencia				
-4,665	2,37112	,24075	-5,14345	-4,18768	-19,379	96	,000

Tabla 10. Descriptivos (tasa de natalidad).

		Media	Desviación típica	Error típico	I. confianza media 95%		Mínimo	Máximo
Europa Oriental	11	14,76	3,69	1,11	12,28	17,24	11,60	24,70
Sudamérica	12	29,17	7,38	2,13	24,48	33,86	18,00	46,60
Europa Occidental	19	12,85	1,94	44,547	11,91	13,78	9,70	16,70
Japón u Oriente Medio	11	33,90	8,62	2,60	28,10	39,69	22,30	45,60
Asia	16	30,00	9,05	2,26	25,1731	34,82	11,70	42,20
África	27	44,52	5,68	1,09	42,2767	46,77	31,10	52,20
Total	96	29,28	13,60	1,38	26,5330	32,04	9,70	52,20

Tabla 11. ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	13964,322	5	2792,864	69,441	,000
Intra-grupos	3619,747	90	40,219		
Total	17584,070	95			

Algunas dificultades y errores previsibles

Lectura de gráficos

En proyectos anteriores hemos analizado algunos niveles de lectura de los gráficos. Cuando los niveles de lectura de gráficos descritos se aplican no sólo a su interpretación, sino a su valoración crítica, los niveles superiores descritos antes se modifican ligeramente (Aoyama y Stephen, 2003; Aoyama, 2007).

Supongamos, por ejemplo, que se da a los estudiantes un gráfico que presenta datos sobre el número de horas que los adolescentes dedican a jugar con la videoconsola y el número de episodios de violencia escolar en que se ven implicados. Dicha gráfica muestra claramente un crecimiento del número de episodios de violencia cuando aumenta el tiempo dedicado a este tipo de juegos. Se pregunta a los estudiantes si piensan que la violencia escolar disminuiría si se prohibiesen las videoconsolas. Una vez que los estudiantes llegan a la fase superior en la clasificación de Friel, Curcio y Bright, todavía podríamos diferenciar tres grupos, en función de su capacidad de crítica de la información representada:

- *Nivel racional/literal*. Los estudiantes leen correctamente el gráfico, incluyendo la interpolación, detección de tendencias y predicción. Para responder la pregunta planteada, usan las características del gráfico, pero no cuestionan la información, ni dan explicaciones alternativas. Una respuesta típica a la pregunta planteada sería "Sí, ya que el grupo de chicos que jugó juegos durante mucho tiempo también tuvo muchos episodios de violencia"
- *Nivel crítico*. Los estudiantes leen los gráficos, comprenden el contexto y evalúan la fiabilidad de la información, cuestionándola a veces, pero no son capaces de buscar otras hipótesis: "Pienso que no, pues aunque los que más juegan aparecen como más violentos en el gráfico, podría haber otras causas, aunque no me imagino cuáles"
- *Nivel hipotético*: Los estudiantes leen los gráficos, los interpretan y evalúan la información, formando sus propias hipótesis y modelos: "No estoy de acuerdo en que la causa de la violencia sea el juego; quizás la falta de atención de los padres puede llevar a la vez a que el chico sea violento y que dedique más horas a jugar con la consola".

Medidas de posición central

En este proyecto los datos originales son medias ponderadas, tema en que los estudiantes tienen dificultades. Pollatsek, Lima y Well (1981) encontraron que incluso alumnos universitarios no ponderan adecuadamente los valores y en ocasiones usan la media simple, en lugar de la media ponderada.

Li y Shen (1992) indican que cuando se pide a los estudiantes calcular la media a partir de una tabla de frecuencias donde los datos se agrupan en intervalos, los estudiantes olvidan con frecuencia que cada uno de estos grupos debe ponderarse de modo distinto al calcular la media.

En otros casos el algoritmo se aplica de forma mecánica sin comprender su significado. Cai (1995) encontró en su investigación que mientras la mayoría de alumnos de 12-13 años eran capaces de aplicar adecuadamente el algoritmo para calcular la media, sólo algunos eran capaces de determinar un valor desconocido en un conjunto pequeño de datos para obtener un valor medio dado. Incluso, encontrando el valor desconocido, fueron pocos los que lo hicieron a partir de un uso comprensivo del algoritmo, multiplicando el valor medio por el número de valores para hallar la suma total. La mayoría simplemente usó el ensayo y error.

Otros errores de cálculo en media, mediana y moda descritos por Carvalho (1998) al analizar las producciones escritas de los alumnos al resolver tareas estadísticas son los siguientes:

- Moda: tomar la mayor frecuencia absoluta.
- Mediana: no ordenar los datos para calcular la mediana; calcular el dato central de las frecuencias absolutas ordenadas de forma creciente; calcular la moda en vez de la mediana; equivocarse al calcular el valor central.
- Media: hallar la media de los valores de las frecuencias; no tener en cuenta la frecuencia absoluta de cada valor en el cálculo de la media.

En realidad, el cálculo de la mediana es complejo, porque el algoritmo de cálculo es diferente, según tengamos un número par o impar de datos, y según los datos se presenten en tablas de valores agrupados o sin agrupar (Cobo y Batanero, 2000) y también el valor obtenido es diferente, según se aplique uno u otro algoritmo. Esto puede resultar difícil para los alumnos que están acostumbrados a un único método de cálculo y una única solución para los problemas matemáticos.

Gattuso y Mary (1998, 2002) analizan la evolución de la comprensión del algoritmo de cálculo de la media ponderada de los alumnos durante la enseñanza secundaria, usando problemas con diferentes contextos y forma de representación. Las tareas presentadas fueron: cálculo de medias ponderadas, efecto que el cambio de un dato produce sobre la media y hallar un valor faltante en un conjunto de datos para obtener un promedio dado. Identifican las siguientes variables didácticas que afectan a la dificultad de las tareas: formato (tabla, serie de números, gráfico), si los valores de las variables son o no mucho mayores que los de las frecuencias (lo que influye en que el niño discrimine los dos conceptos); si una de las frecuencias es mucho mayor que las otras (de modo que se fuerce al niño a tener en cuenta las frecuencias). Observaron el efecto de estas variables y también la mejora con la instrucción, aunque no fue muy persistente en el tiempo.

Correlación y regresión

La asociación estadística extiende la idea de dependencia funcional, y es relevante en muchos métodos de gran variedad de ciencias. Mientras que en una dependencia de tipo funcional a cada valor de una variable independiente X corresponde un solo valor de otra variable dependiente Y, en el caso de asociación a cada valor de X corresponde una distribución de valores de Y. Además se puede definir una medida de la intensidad de la asociación, que varía entre 0 (independencia total) y 1 (asociación perfecta), utilizando coeficientes como el de correlación (para variables cuantitativas) o el coeficiente de contingencia (para variables cualitativas).

Las dificultades y errores en torno a estas nociones, descritas por la investigación psicológica, están ligadas a las ideas de condicionamiento y causalidad, aunque estos conceptos son diferentes. Díaz y de la Fuente (2005) analizan las relaciones entre causalidad y condicionamiento, que son inherentes a un problema de asociación, puesto que las personas construimos nuestro conocimiento del mundo sobre la base de relaciones de causa y efecto entre diferentes sucesos.

Desde el punto de vista de la probabilidad, si un suceso A es la causa estricta de un suceso B, siempre que suceda A, sucederá B, por lo que P(B/A) = 1 y habrá una asociación entre A y B. Las autoras indican que una relación causal determinista es menos frecuente que una aleatoria, donde, cuando sucede A cambia la probabilidad de que ocurra B. Una relación de causalidad implica una asociación, pero no al contrario, pues dos sucesos o variables pueden estar asociados sin que uno de ellos sea causa del otro. Una asociación estadística entre variables puede deberse a otras variables intervinientes o incluso ser espúrea y no implica relación causal.

Otros sesgos descritos relacionados con la comprensión de la idea de correlación citados por Engel y Sedlmeier (2011) son los siguientes:

- *Influencia de las creencias previas o correlación ilusoria*, que consiste en formarse teorías propias sobre la relación que existe entre dos variables, que no se apoya en los datos empíricos.
- *Desestimar el efecto de regresión*, por el cual, después de que aparece un caso atípico, suele haber una vuelta a los valores más comunes. Así, los hijos de padres muy altos suelen ser altos, pero no tanto como sus padres.
- *No tener en cuenta* el efecto de terceras variables que pueden hacer creer en la existencia de una correlación cuando ésta no existe.

En la investigación en didáctica de la matemática se han encontrado las siguientes dificultades adicionales:

- *Distinguir* si las dos variables presentadas en la tarea conforman una distribución bidimensional (Estepa, 2007).

- La mayoría de los alumnos *relaciona el signo de la covarianza* con el tipo de correlación directa o inversa, y se tiene dificultad en relacionar su magnitud con la intensidad o relacionar los coeficientes de regresión con el signo de la correlación (Sánchez-Cobo 1999; Estepa, 2007).
- Considerar tan sólo la correspondencia de datos bidimensionales aislados, en lugar de considerar la tendencia global de éstos (Moritz, 2004).
- Estimar el coeficiente de correlación desde otras representaciones (verbal tabular, etc.) distintas de la representación gráfica (diagrama de dispersión). En este sentido, destacar la importancia del trabajo con distintos objetos matemáticos cuya comprensión y dominio permitirán llevar a cabo procesos de traducción entre diferentes representaciones de la asociación estadística (verbal, tabular, gráfica o numérica) (Sánchez Cobo, 1999; Moritz, 2004).
- Ordenar adecuadamente diferentes valores del coeficiente de correlación distintos de: –1, 0 y 1. Aunque tienen presente que la intensidad de la dependencia se obtiene a partir del coeficiente de correlación, presentan dificultades en estas comparaciones, debido a que usan indebidamente sus conocimientos sobre el orden en números enteros (Sánchez-Cobo, 1999).
- Dificultad de diferenciar la variable explicativa de la explicada en el cálculo de la recta de regresión. En la investigación de Sánchez-Cobo, dos de cada cinco estudiantes, aproximadamente, relacionan que ambas rectas de regresión son perpendiculares cuando el coeficiente de correlación es nulo, evidenciándose que se considera casi en exclusiva, la modelización de ajuste lineal. En concreto, casi la mitad de los sujetos del estudio consideran que si existe correlación positiva, ésta se deduce en una dependencia lineal.
- Transitividad de la correlación. Se piensa que si una variable A está correlacionada con otra B y ésta se relaciona con una variable C, las variables A y C debieran estar correlacionadas, pero en general, la correlación no es transitiva. Esto se debe a una extrapolación de propiedades que se cumplen en otras relaciones (como la de orden). Esta creencia es denominada concepción transitiva de la asociación (Sotos, Vanhoof, Van Den Noortgate y Onghena, 2009).

Por otra parte, el estudio de la regresión ofrece al alumno una buena ocasión de aplicar sus conocimientos previos sobre la función lineal y otras funciones sencillas. El profesor debe advertir sobre el uso inadecuado del razonamiento proporcional en el trabajo con la recta de regresión (Estepa, 2008) y recordar que, en general, no se puede despejar la variable independiente para encontrar la línea de regresión de X sobre Y, porque el criterio de ajuste es diferente del seguido en la construcción de la recta de regresión de Y sobre X.

Otras dificultades

En este fichero las unidades estadísticas son agregados, lo que tiene repercusión en la interpretación de los promedios de las variables; por ejemplo, la media (simple) de todas las esperanzas de vida al nacer en los distintos países no es igual a la esperanza global de vida, sino que ésta tiene que ser calculada como una media ponderada en función del número de hombres/mujeres de cada país. Esto puede servir de reflexión sobre las diferentes unidades estadísticas que pueden usarse en un estudio, la necesidad que a veces tenemos de trabajar con valores aproximados y la forma de combinar estudios parciales para obtener índices globales, así como para reflexionar sobre las propiedades de los promedios y la comprensión de las mismas por parte de los estudiantes (Tormo, 1995; Batanero, 2000). Esta misma observación debe hacerse en la interpretación de otros estadísticos, como los de dispersión o los coeficientes de correlación.

El trabajo con análisis exploratorio de datos refuerza algunos objetivos sugeridos para la educación matemática en los nuevos currículos de secundaria, como el trabajo con problemas abiertos, el uso de sistemas múltiples de representación, la introducción al trabajo con ordenadores o calculadoras gráficas y la conexión de las matemáticas con otras áreas del currículo (Shaughnessy, Garfield y Greer, 1997).

Pero el razonamiento estadístico va más allá del conocimiento matemático y de la comprensión de los conceptos y procedimientos. La modelización, la valoración de la bondad del ajuste de los modelos a la realidad, la formulación de cuestiones, la interpretación y síntesis de los resultados, la elaboración de informes son también componentes esenciales de las capacidades que queremos desarrollar en nuestros alumnos. Los estudiantes, acostumbrados a que en la clase de matemáticas cada problema tenga una única solución, podrían encontrar complejo el trabajo con los proyectos y la existencia de diferentes procedimientos y soluciones correctas. Es labor del profesor acostumbrarles al método y razonamiento estadístico.

Los estadísticos de orden pueden plantear problemas, ya que la función de distribución de la variable discreta tiene discontinuidades de salto y los alumnos no están acostumbrados a este tipo de funciones. Asimismo, alguno de ellos pudiera tener deficiencias en el razonamiento proporcional.

Análisis del contenido estadístico

En este proyecto podemos identificar, explícita o implícitamente, los siguientes contenidos:

1. Aplicaciones de la estadística:

- Estudios demográficos y socioeconómicos.
- Estadísticas oficiales; organismos y procedimientos en la elaboración de estadísticas oficiales.
- Fuentes de datos estadísticos: anuarios estadísticos.
- fuentes de datos en Internet.
- Ajuste de modelos a datos.

2. Conceptos y propiedades:

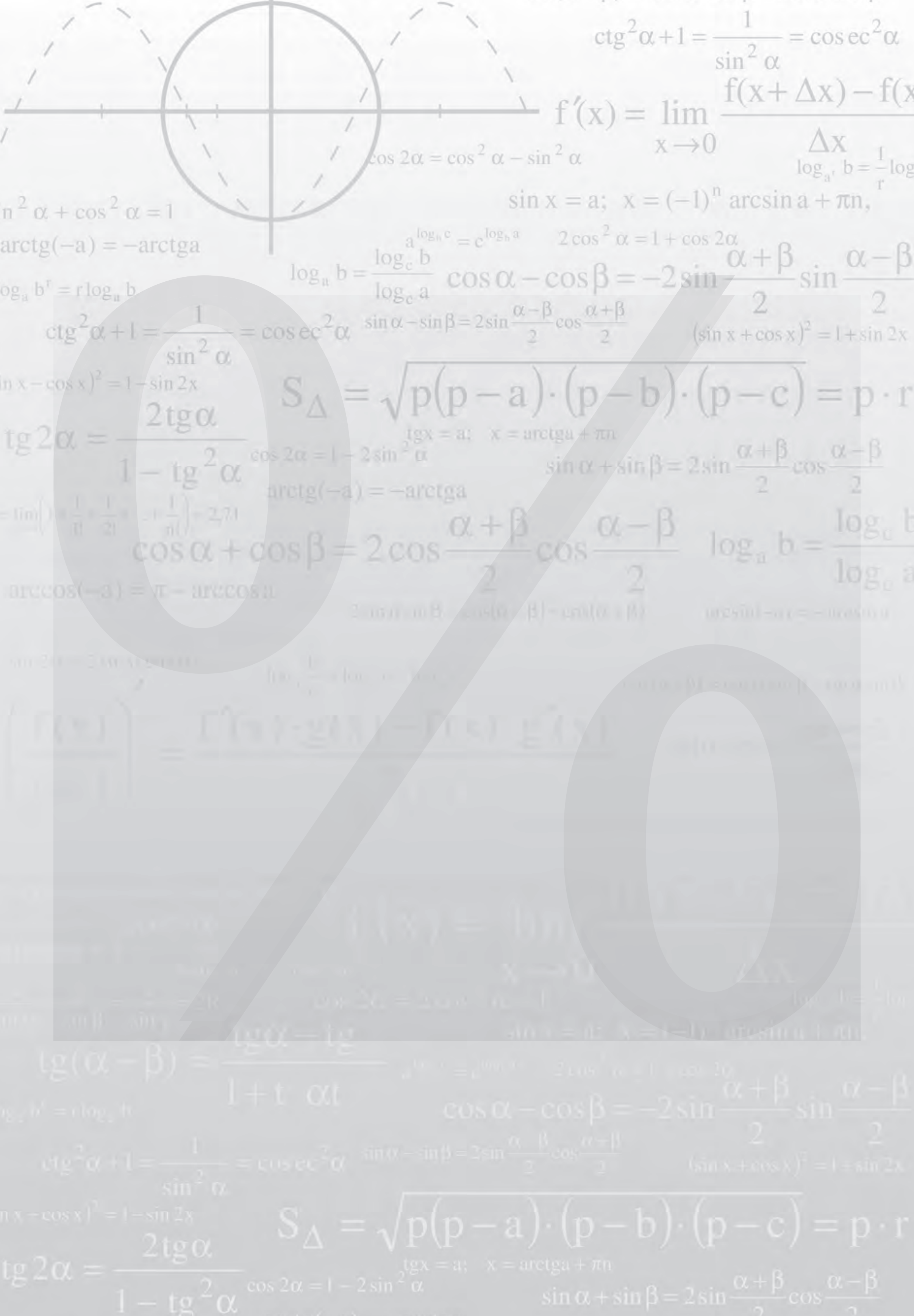
- Agrupación de variables en intervalos, efectos de la agrupación.
- Medidas de posición central: medias ponderadas, percentiles, rangos de percentiles.
- Valores atípicos y su efecto sobre los promedios.
- Asociación y sus tipos: directa/inversa; lineal/no lineal; asociación y causalidad.
- Correlación. Interpretación del signo y la intensidad, proporción de varianza explicada.
- Modelos: ajuste de modelos sencillos a datos bivariantes; uso de modelos en la predicción. Función lineal, cuadrática, logarítmica, exponencial.
- Contraste de hipótesis: contraste de diferencia de medias en muestras relacionadas.
- Intervalo de confianza para la diferencia de medias en muestras relacionadas.
- Análisis de varianza de un factor, efectos fijos. Variable dependiente y factor.

3. Notaciones y representaciones:

- Palabras como intervalos, extremos y marcas de clase, percentiles y sus rangos, cuartiles, valores atípicos, asociación, correlación, regresión, etc.

- Símbolos usados para los diferentes estadísticos y otros como el sumatorio.
- Gráficos de barras, diagramas acumulativos, diagrama de tallo y hojas, gráficos de caja, gráficos de cuantiles, diagramas de dispersión.

4. *Técnicas y procedimientos:*
    - Búsqueda de datos a partir de anuarios estadísticos o de Internet.
    - Elaboración de tablas de frecuencia con datos agrupados, desarrollo de gráficos de tallo y hojas, gráficos de caja, gráficos de cuantiles y diagramas de dispersión.
    - Interpretación de tablas y gráficos: elaboración de conclusiones a partir del análisis de tablas y gráficos.
    - Elaboración de argumentos y conclusiones a partir del análisis de datos.
    - Estimación de parámetros en modelos de regresión.
    - Uso de calculadora gráfica, hojas de cálculo o software estadístico.
    - Comparación de muestras relacionadas: contrastes e intervalos de confianza.
    - Análisis de varianza, cálculo e interpretación.
5. Actitudes:
    - Valorar la utilidad y complejidad de la elaboración de las estadísticas oficiales y resaltar la importancia de su colaboración en encuestas y censos para obtener datos fiables.
    - Concientizar al alumno sobre la importancia de las representaciones gráficas y la posibilidad de que se transmita información sesgada en una gráfica mal construida.
    - Fomentar un espíritu crítico en el uso de paquetes estadísticos y sus opciones por defecto.



$\mathrm{ctg}^2\alpha+1=\frac{1}{\sin^2\alpha}=\mathrm{cosec}^2\alpha$
$f'(x)=\lim\limits_{x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$
$\cos 2\alpha=\cos^2\alpha-\sin^2\alpha$
$\sin x=a;\ \ x=(-1)^n\arcsin a+\pi n,$
$\sin^2\alpha+\cos^2\alpha=1$
$\mathrm{arctg}(-a)=-\mathrm{arctga}$
$a^{\log_b c}=c^{\log_b a}$
$2\cos^2\alpha=1+\cos 2\alpha$
$\log_a b=\frac{\log_c b}{\log_c a}$
$\cos\alpha-\cos\beta=-2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$
$\log_a b^r=r\log_a b$
$\sin\alpha-\sin\beta=2\sin\frac{\alpha-\beta}{2}\cos\frac{\alpha+\beta}{2}$
$(\sin x+\cos x)^2=1+\sin 2x$
$(\sin x-\cos x)^2=1-\sin 2x$
$S_\Delta=\sqrt{p(p-a)\cdot(p-b)\cdot(p-c)}=p\cdot r$
$\mathrm{tg}\,2\alpha=\frac{2\mathrm{tg}\alpha}{1-\mathrm{tg}^2\alpha}$
$\mathrm{tg}x=a;\ \ x=\mathrm{arctga}+\pi n$
$\cos 2\alpha=1-2\sin^2\alpha$
$\sin\alpha+\sin\beta=2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$
$\cos\alpha+\cos\beta=2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$
$\arccos(-a)=\pi-\arccos a$
%

Semblanza del Dr. Ernesto Alonso Sánchez Sánchez

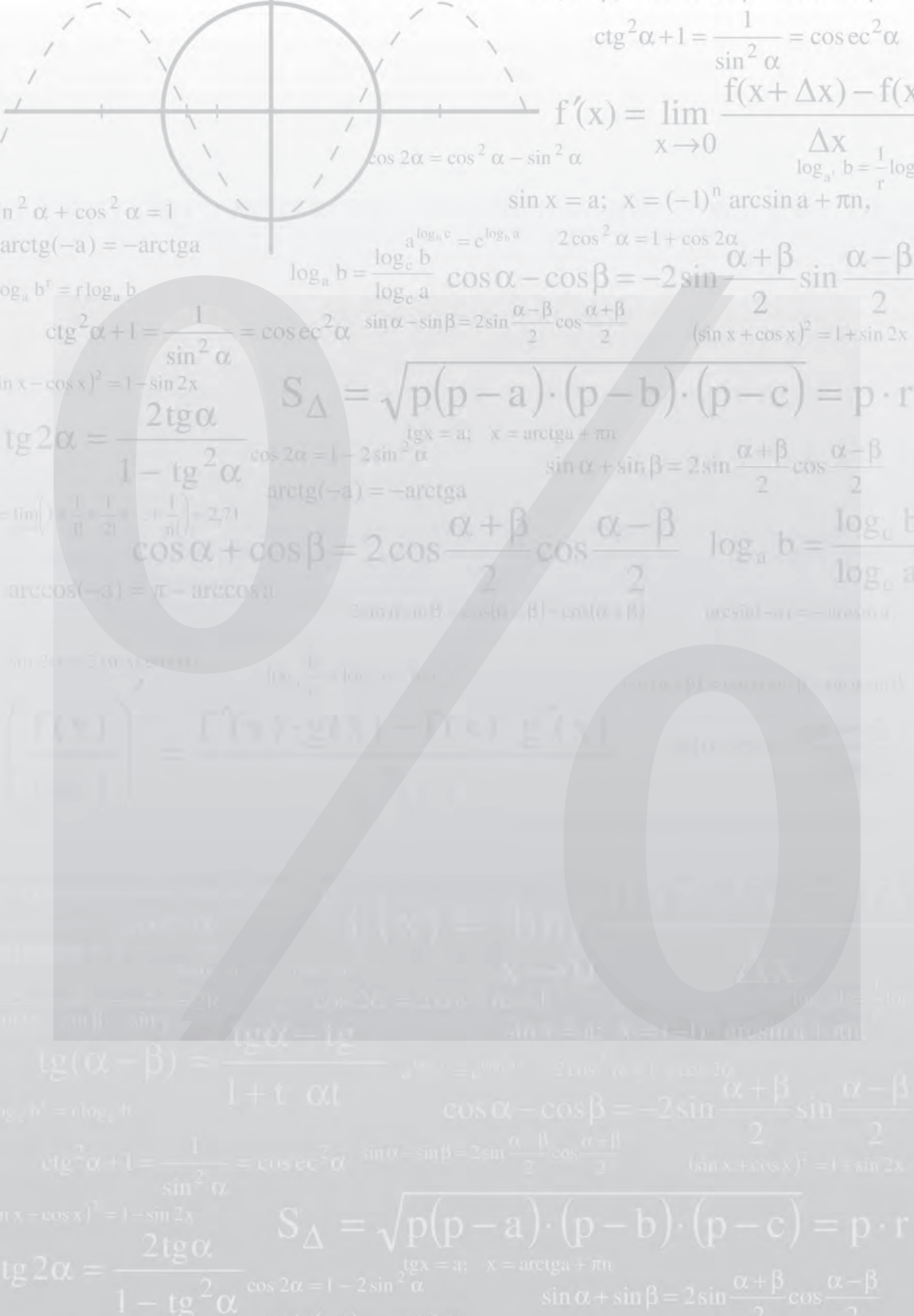
Licenciado en Matemáticas, por la Facultad de Ciencias de la UNAM, y obtuvo su Maestría y Doctorado en Ciencias, en la especialidad de Matemática Educativa, por el Cinvestav. Hizo una estancia posdoctoral en Estrasburgo, Francia, (1997-98) y estancias sabáticas en Portland, Estados Unidos, (2003-04) y Granada, España, (2011-12). Es miembro del SNI, nivel 1. Ha dirigido ocho tesis de doctorado, 28 tesis de maestría. Es editor asociado de la Statistical Education Research Journal (SERJ) y lo fue de Innovación Educativa. Ha sido *referee* en varias revistas, entre las que se encuentran Educación Matemática, Revista Latinoamericana de Investigación en Matemática Educativa, Publicaciones, Innovación Educativa, Journal of Mathematicas Teacher Education, SERJ. Ha publicado 16 artículos en revistas internacionales o capítulos de libros y más de 38 artículos en extenso en congresos internacionales. Su línea de investigación es la didáctica de la probabilidad y la estadística, en la que ha hecho investigaciones sobre el concepto de variación (Conacyt, proyecto), distribución (Conacyt) y competencias estadísticas. Es editor del libro *Aprendizaje y enseñanza de las matemáticas escolares*, publicado por la SEP y destinado a los profesores de educación básica del país. Entre sus más recientes publicaciones se encuentran:

Landín, P. y Sánchez, E. (2010). Niveles de razonamiento probabilístico de estudiantes de bachillerato frente a tareas de distribución binomial. Educación Matemática Pesquissa, São Paulo, 12(3), pp. 598-618.

Sánchez, E. (2010). "Una jerarquía de razonamiento estadístico sobre la noción de predicción/incertidumbre elaborada con profesores de secundaria". *Revista Latinoamericana de Investigación en Matemática Educativa*, número especial. vol. 13(4): 129-158.

Sánchez, E. , Borim, C. & Coutinho, C. (2011). "Teachers' Understanding of Variation". En C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching* *and Teacher Education. A joint ICMI-IASE Study (211-221)*. Nueva York, EU: Springer.

Sánchez, E. y Gómez, A.L. (2011) "El desarrollo del pensamiento estadístico de profesores de secundaria en servicio." En Juan J. Ortiz, (Ed.). *Investigaciones actuales en educación estadística y formación de profesores,* (55-72). Granada, España.


$ctg^2\alpha + 1 = \frac{1}{\sin^2\alpha} = \cos ec^2\alpha$
$f'(x) = \lim_{x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$
$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$
$\sin x = a; \quad x = (-1)^n \arcsin a + \pi n,$
$\sin^2\alpha + \cos^2\alpha = 1$
$arctg(-a) = -arctga$
$a^{\log_b c} = c^{\log_b a}$
$2\cos^2\alpha = 1 + \cos 2\alpha$
$\log_a b = \frac{\log_c b}{\log_c a}$
$\cos\alpha - \cos\beta = -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$
$\log_a b^r = r\log_a b$
$\sin\alpha - \sin\beta = 2\sin\frac{\alpha-\beta}{2}\cos\frac{\alpha+\beta}{2}$
$(\sin x + \cos x)^2 = 1 + \sin 2x$
$(\sin x - \cos x)^2 = 1 - \sin 2x$
$S_\Delta = \sqrt{p(p-a)\cdot(p-b)\cdot(p-c)} = p \cdot r$
$tg 2\alpha = \frac{2tg\alpha}{1 - tg^2\alpha}$
$tgx = a; \quad x = arctga + \pi n$
$\cos 2\alpha = 1 - 2\sin^2\alpha$
$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$
$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$
$\arccos(-a) = \pi - \arccos a$

BIBLIOGRAFÍA

Aoyama, K. (2007), Investigating a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education 2*(3). Disponible en: www.iejme/.

Aoyama, K. y Stephens, M. (2003), Graph interpretation aspects of statistical literacy: A Japanese perspective, *Mathematics Education Research Journal 15*(3), 3-22.

Arteaga, Batanero, Ortiz y Contreras (2011).

Batanero, C. (2000a), Controversies around significance tests,

Mathematical Thinking and Learning, 2(1-2), 75-98.

Batanero, B. (2000b), Significado y comprensión de las medidas de posición central. *UNO*, 25, 41-58.

Batanero, C. y Díaz, C. (Eds.), (2011). *Estadística con proyectos*, Granada, Departamento de Didáctica de la Matemática.

Batanero, C. y Sánchez, E. (2005), What is the nature of high school student's conceptions and misconceptions about proba- bility? En G. Jones (Ed.), *Exploring probability in school: challeng- es for teaching and learning* Nueva York: Springer. (pp. 241-266).

Beneitone, P., Esquetini, C., González, J., Maletá, M. M., Suifi, G. y Wagenaar, R. (Eds.) (2007). *Reflexiones y Perspectivas de la Educación Superior en América Latina*. Informe final – Proyecto Tuning – América Latina 2004-2007, Bilbao: Publicaciones de la Universidad de Deusto.

Burril, G. y Bielher, R. (2011), Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero,

Burril, G., Reading, C. (Eds.), *Teaching Statistics in School Mathematics – Challenges for teaching an teacher education* (Nueva York: Springer. (57-69).

Cai, J. (1995), Beyond the computational algorithm.Students' unsderstanding of the arithmetic average concept. En L. Meira (Ed.), *Proceedings of the 19th PME Conference* (Vol. 3, pp. 144-151). Universidade Federal de Pernambuco, Recife, Brazil.

Carvalho, C. (1998), *Tarefas estadísticas e estratégias de resposta*.

Trabajo presentado en el VI Encuentro en Educación Matemática de la Sociedad Portuguesa de Ciencias de la Educación. Castelo de Vide, Portugal.

Castro-Sotos, A. E., Van Hoof, S., Van den Noortgate, W. y Onghena, P. (2009). The transitivity misconception of Pearson's correlation coefficient. *Statistics Education Research Journal, 8*(2), 33-55.

Cobb, P. & Moore, D. (2000), Statistics and mathematics: ten- sion and cooperation. The American Mathamtical Monthly (agosto-septiembre), pp.615-630.

Cobo, B. y Batanero, C. (2000), *La mediana en la educación secundaria obligatoria: ¿Un concepto sencillo?* UNO 23, pp. 85-96.

Curcio, F. R. (1987). Comprehension of mathematical relation- ships expressed in graphs. *Journal for Research in Mathematics Education*, 18 (5), 382-393.

Day, A. (Ed.) (1992), *The annual register 1992*, 234. London: Longmans.

Díaz, C., Arteaga, P. y Batanero, C. (2008), Contribución del trabajo con proyectos estadísticos a la adquisición de competencias básicas. En M., Molina, P. Perez-Tyteca y M. Fresno (2008). *Investigación en el aula de matemáticas*. *Competencias matemáticas*. Granada: SEAM. Thales y Departamento de Didáctica de la Matemática de la Universidad de Granada.

Díaz, C. y De la Fuente, I. (2005), Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Epsilon, pp. 59*. 245-260.

Díaz-Barriga, F. (2012). Reformas curriculares y cambio sistémico: una articulación ausente pero necesaria para la innovación. *Revista Iberoamericana de Educación Superior (RIES), 3*(7), 23-40.

Estepa, A. (1995). *Consideraciones sobre la enseñanza de la asociación estadística. UNO, 5, 69-79.*

Estepa, A. (2007). Caracterización del significado de la correlación y regresión en estudiantes de educación secundaria. *Zetetiké* 15(28), 119-151. Disponible en: http://ries.universia.net/index.php/ries/article/view/229]

Fischhoff, B. y Kadvany, J. (2011), Risk: *A very short introduction*. EU: Oxford University Press.

Franklin, C., Kader, G., Newborn, D. S., Moreno, J., Peck, R., Perry, M. & Schaeffer, R. (2005), A Curriculum Framework for pre K-12 *Statistics Education. Informe presentado a la American Statistical Association*.

Gal, I. (2004). Statistical literacy: Meanings, componentes, responsabilities. En D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking (47-78)*. Dordrecht: Kluwer.

Gattuso, L. y Mary, C. (1996). Development of concepts of the arithmetic average from high school to University. Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education. (Vol. I, pp. 401-408). Universidad de Valencia.

Gattuso, L. y Mary, C. 2002). Development of the concept of weighted average among high-school children. En B. Phillips (Ed.), Proceedings of the Sixth International Conference on Teaching Statistics. [CD-ROM]. Cape Town: International Asso- ciation forStatistical Education.

Goñi Zabala, J.M. (2005), *El espacio europeo de educación superior, un reto para la universidad*. Barcelona: Octaedro / ICE Universidad de Barcelona.

Goñi Zabala, J.M. (2008). 3^2-2 *ideas clave. El desarrollo de la competencia matemática*. Barcelona: Editorial GRAÓ.

INCE (2004). *Marcos teóricos de PISA 2003: la medida de conocimientos y destrezas en matemáticas, lectura, ciencias y resolución de problemas*. Madrid: MEC. Disponible en: www.ince.mec.es/ pub/marcoteoricopisa2003.pdf

Li, D. Y. y Shen, S. M. (1992). Students' weaknesses in statistical projects. *Teaching Statistics* 14 (1), 2-8.

MacGillivray, H. & Pereira-Mendoza, L. (2011). Teaching statis- tical thinking through investigative projects. En D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking (109-120)*. Dordrecht: Kluwer.

Mayén, S. (2009). *Comprensión de las medidas de tendencia central en estudiantes mexicanos de educación secundaria y bachillerato*, tesis doctoral no publicada, Universidad de Granada.

Monteiro, C. y Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers, en *International Electronic Journal of Mathematics Education* 2(3),188-207. Disponible en: http://www.iejme/.

Moritz, J. (2004). Reasoning about covariation. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 221-255). Dordrecht: Kluwer.

Moore, D. S. (1998). Statistics among the liberal arts. *Journal of American Statistical Association*, 93, 1253-1259.

Moore, D. S. (1992). Teaching statistics as a respectable subjetct. En F. Gordon y S. Gordon (Eds.), *Statistics for the Twenty-first century (14-25)*. Washington D.C.: Mathematical Association of America.

Moore, D. S. NCTM (2000). *Principles and Standards for School Mathematics*. Reston, VA.: The National Council of Teachers of Mathematics.

Orta, J.A. (2009). *Estudio exploratorio sobre comparaciones de conjuntos de datos con estudiantes de secundaria*. Tesis de maestría no publicada. Departamento de Matemáticas Educativa.

Pollatsek, A., Lima, S. y Well, A.D. (1981). Concept or computation: *Students'understanding of the mean. Educational Studies in Mathematics, 12*,191-204.

Rossman, A., Chance, B. & Medina, E. (2006). Some important comparisons between statistics and mathematics, and why teachers should care. En G.F. Burril y P.C. Elliot (Eds.), *Thinking and Reasoning with Data and Chance. (Sixty-eight Yearbook]*. Reston, VA.: National Council of Teachers of Mathematics.

Perrenoud, P. (2002). *Construir competencias en la escuela*, Santiago de Chile: Ed. Dolmen.

Rouncenfield (1995). The statistics of poverty and inequality, *Journal of Statistics Education, 3*(2).

Sánchez Cobo, F.T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral. Universidad de Granada.

Schield, M. (2006). Statistical literay survey analysis: reading graphs and tables of rates percentages. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town: International Association for Statistical Education. Disponible en: http://www. stat.auckland.ac.nz/~iase.

Schoenfeld, A. & Hermann, D. (1982). Problem perception and knoeledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory and Cognition 8*(5), 484-494.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data andchance. In F. Bidulph& K. Carr (eds.) *Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australasia*, 6-22. Rotorua, N.Z.: Universidad of Waikata.

TORMO, C. (1995). *Dificultades del alumnado respecto a la media aritmética*, UNO, 5, 29-36.

UNESCO. (1990). *Demographic year book 1990*. Nueva York: Naciones Unidas.

WATSON, J.M. (2006). *Statistical Literacy at School: Growth and Goals*. Mahawah, NJ.: Lawrence Erlbaum Associates, Publishers.

WATSON, J. M. y Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*(1 y 2), 11-50.

WILD, C. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-262.

Se terminó de imprimir y encuadernar en diciembre de 2013
en Impresora y Encuadernadora Progreso, S. A. de C. V. (IEPSA),
Calzada San Lorenzo 244; C.P. 09830, México, D. F.
El tiraje fue de 10,000 ejemplares.

ISBN: 978-607-9362-00-3
9 786079 362003

SEP
SECRETARÍA DE EDUCACIÓN PÚBLICA

ESTADOS UNIDOS MEXICANOS

SUBSECRETARÍA DE EDUCACIÓN MEDIA SUPERIOR www.sems.gob.mx