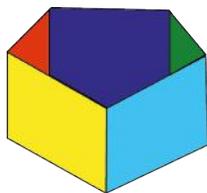

Reforma de la Educación Matemática en Costa Rica



www.reformamatematica.net



La enseñanza de la Estadística:
más allá de procedimientos y técnicas

Unidad didáctica

Primera Parte



Curso bimodal de capacitación para docentes de la
Educación Secundaria
2015

Versión colocada en Plataforma el 07 de octubre de 2015.

Contenidos

Introducción.....	2
Orígenes de la Estadística y su enseñanza	3
Alfabetización en Estadística	9
Recolección de datos.....	12
Análisis didáctico.....	17
Muestreo estadístico e inferencia estadística	20
Análisis didáctico.....	31
Elaboración de bases de datos	35
Análisis didáctico.....	37
Bibliografía.....	44
Créditos.	45

Introducción

La presente unidad didáctica ha sido elaborada con el propósito de apoyar la labor de los docentes de Matemáticas en su quehacer educativo específicamente en lo referente a la enseñanza de la Estadística en la educación secundaria. La discusión que se realiza está en concordancia con lo establecido en los Programas de Estudio de Matemáticas del Ministerio de Educación Pública.

La enseñanza de esta disciplina, en el mejor de los casos, ha estado dirigida a la implementación de procedimientos y técnicas para la recolección, procesamiento y presentación de información. Sin embargo, muchas veces se descuidan pilares disciplinares que trascienden las fórmulas y mecanismos algorítmicos, por lo que se pierde de vista el propósito básico que llevó a hacer necesarios estos instrumentos. La Estadística es mucho más que tablas, gráficas, medidas o procedimientos algorítmicos; es una disciplina científica cuyo propósito más elemental consiste en describir, interpretar y poner en contexto los patrones de variabilidad de la información que se generan alrededor de diferentes áreas científicas. La enseñanza de la Estadística debe tomar en cuenta este propósito para motivar su aprendizaje, de manera que permita a los estudiantes adquirir las habilidades necesarias para comprender fenómenos de diferente naturaleza, los cuales se vinculan con el análisis e interpretación de datos.

Producto de lo anterior, la discusión se remonta al análisis, interpretación y lectura de la información que puede ser comunicada mediante las diferentes técnicas estadísticas incluidas en los programas de estudio, en procura de dar respuesta a problemas que involucran datos tanto cuantitativos como cualitativos. Las actividades que aquí se presentan están dirigidas al docente de Matemáticas; en estas se discuten aspectos didácticos muy propios de la enseñanza de la disciplina, que el docente debe tomar en cuenta al realizar su labor.

Por razones propias del curso de capacitación que imparte el Ministerio de Educación Pública, la unidad didáctica ha sido dividida en dos documentos; no obstante, se solicita al lector realizar un análisis integral de su contenido.

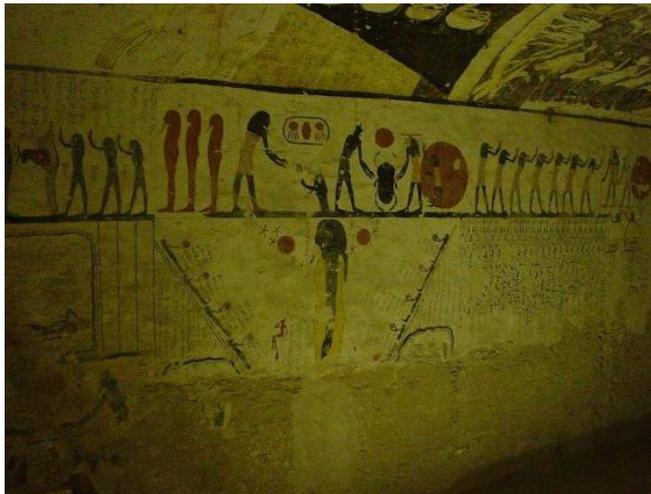
Orígenes de la Estadística y su enseñanza

La Estadística es una disciplina que tiene sus orígenes hace miles de años, cuando el ser humano requirió de establecer estrategias para sistematizar información que fuera de utilidad para sus actividades cotidianas. Se han encontrado en cuevas diversos dibujos de objetos y animales, que constituyen formas artesanales de representación de la información que encontraba en su entorno.



Imágenes con derechos adquiridos por el MEP

Según estudiosos, muchos de estos dibujos se utilizaban para llevar la cuenta del ganado y la caza. Del mismo modo, se han encontrado tablillas de arcilla con más de cinco mil años de antigüedad, en las que los babilónicos recopilaban datos sobre la producción agrícola y el comercio. Por su parte, los egipcios llevaban cuenta de los movimientos poblacionales mediante la aplicación de censos periódicos. Incluso se habla de la existencia de una diosa de los libros y las cuentas, la diosa Safnkit.



<http://antecedentes.net/?s=antecedentes+estadistica>

Esta necesidad de buscar formas de organizar y resumir la información fue la que dio origen a la Estadística. Sin embargo, varios historiadores consideran que el primer estudio estadístico formal se realizó hasta el año 1662, cuando el inglés John Graunt publicó un tratado con observaciones políticas y naturales incluyendo cifras de nacimientos y defunciones ocurridas en Londres durante el periodo 1604-1661 y su relación con elementos sociales y políticos que afectaban a Londres en ese momento. A partir de ese momento se generó un proceso de desarrollo de la disciplina que se mantiene hasta el tiempo actual, ya que la Estadística es un área en proceso de evolución permanente.

La importancia de la disciplina para los procesos de recolección, organización, clasificación, presentación y análisis de información cuantitativa o cualitativa proveniente de diferentes áreas científicas, dio pie para que la Estadística haya sido considerada una herramienta fundamental para el desarrollo de dichas áreas. Por esta

razón, su enseñanza ha cobrado especial interés no solamente dentro de los currículos universitarios sino también en la enseñanza general básica.

A pesar de lo anterior, los mayores problemas que enfrenta la orientación y el desarrollo de esta área, aun dentro de las mismas comunidades científicas, tiene su origen en el escaso avance de la enseñanza de esta disciplina, que no ha sido capaz de generar el conocimiento necesario sobre los principios que fundamentan el área y su quehacer. Esto ha provocado que existan muchos mitos a su alrededor, hasta llegar a creer que la Estadística carece de formalidad y puede ser utilizada para “*mentir*” lo cual le imprime un carácter de inseguridad al pensar que los estudios basados en ella son poco confiables. Estas creencias obedecen a la ausencia de una cultura básica sobre la disciplina que provoca que ciertas falacias cobren vigencia.

Seguidamente se presentan algunos problemas sobre la forma en que un mal uso de la disciplina conduce a equivocaciones y falsas interpretaciones.

Actividad 1

Pedro es un jugador empedernido de lotería, sus números preferidos son el 47, 62 y 24. Joaquín, un amigo de Pedro, le recomienda que lea el siguiente reportaje del Periódico La Nación:

HAN SALIDO MÁS VECES EN ÚLTIMOS SEIS AÑOS CON CUALQUIER PREMIO

Los números 03, 61 y 83 son los más premiados en la Lotería Nacional

- 47, 57 y 62 han sido los números menos favorecidos en los sorteos dominicales.
- Otros números que han sido muy premiados son: 05, 06, 49, 64, 69, 76 y 82.

JAIRO VILLEGAS S. | jvillegas@nacion.com

La suerte ha alcanzado más a los números 03, 61 y 83 en los sorteos de la Lotería Nacional realizados desde el 6 de enero del 2002 hasta el domingo 7 de diciembre del 2008. Durante ese lapso hubo unos 29.000 premios y esos numerales son los que más repitieron.

MÁS SOBRE ESTE TEMA

La Nación tomó todos los números (del 00 al 99) para determinar la cantidad de veces en que han salido favorecidos con cualquier premio. El promedio es de 290 veces, pero en el caso del 03 ha salido en 328 ocasiones; el 83, en 327, y el 61, en 326.

Como dato curioso, el 03 ha salido como gordo navideño en dos oportunidades (2005 y 1991) y el 83 fue el mayor en el 2001.

Los sorteos del gordo navideño similares al actual se realizan desde el año de 1960.

Otros números que han resultado muy favorecidos en cualquier premio de la lotería dominical son: 05, 06, 49, 64, 69, 76 y 82.

Este análisis no pronostica el resultado del sorteo de la lotería navideña de la Junta de Protección Social de San José (JPS), que se realizará mañana a las 7 p. m. en Alajuela, sino que solo se trata de un recuento histórico.

La acera de las cifras más alejadas de la suerte está encabezada por el número 47, que apenas salió en 248 ocasiones, entre 29.000 premios rifados.

El 57 y el 62 también han sido muy poco afortunados con menos de 260 apariciones. No obstante, el 62 ha sido gordo navideño en los sorteos de 1963 y del 2004, según los datos de la Junta.

Fuente: http://www.nacion.com/ln_ee/2008/diciembre/13/pais1810144.html

Después de analizar esta información,

1. ¿Debería Pedro seguir jugando los mismos números o esta información le proporciona argumentos para cambiarlos?
2. Justifique estadísticamente su respuesta.

Análisis de la actividad 1

El ejemplo anterior presenta una de las falacias más comunes entre los jugadores de lotería. Se denomina “falacia del jugador” y consiste en creer que un número que ha dejado de salir favorecido como premio principal en los sorteos de lotería aumenta sus probabilidades de que salga favorecido en los siguientes sorteos.

Otras personas, por el contrario, creen que los números que más se han repetido en un determinado período de tiempo tienen más probabilidad de salir premiados.

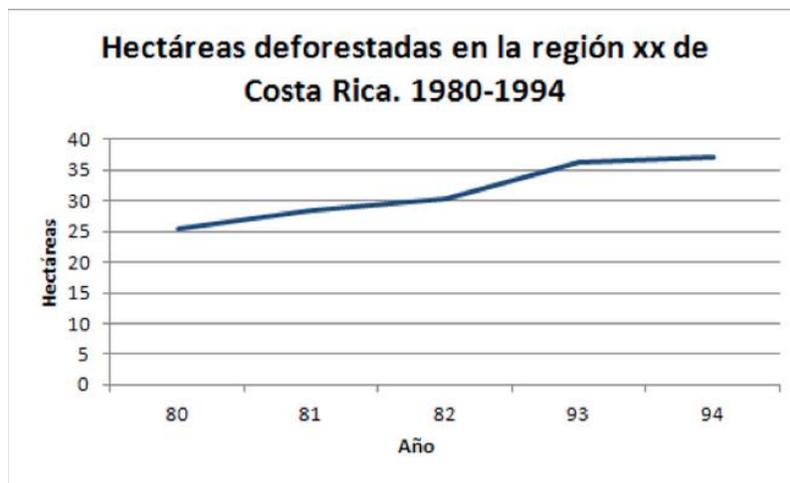
Sin embargo, por el nivel de independencia estadística que tienen los sorteos de lotería, los resultados previos no pueden influir en los sorteos venideros. Por ello, aunque existan números que tienen una alta frecuencia de ocurrencia, esto no les da un mayor o menor favoritismo estadístico. Los sorteos de la Lotería Nacional constituyen un juego de azar en que todos los números participantes tienen la misma probabilidad de ser favorecidos, independientemente de que haya o no haya salido premiado en los sorteos precedentes.

El problema anterior plantea un reto fundamental para el sistema educativo, que consiste en romper con mitos y creencias equivocadas sobre los fenómenos aleatorios. Desde épocas remotas los fenómenos aleatorios han sido vinculados con el misticismo. Por muchos años se ha creído que los eventos aleatorios responden a leyes sobrenaturales, por lo que sus resultados dependen de la voluntad divina. En este sentido el lector puede profundizar en la temática debido a que se ha escrito bastante sobre el tema. Al ilustre físico Albert Einstein se le atribuye la famosa frase “*Dios no juega a los dados con el universo*”, la cual ha alimentado la discusión acerca de la presencia de la aleatoriedad en situaciones cotidianas.

Seguidamente se plantea un problema que complementa el problema anterior en relación con el uso de la Estadística.

Actividad 2

La siguiente representación gráfica fue publicada en un informe de una empresa consultora de Costa Rica.



Fuente: Información suministrada por X.¹

¹ No se incluye el nombre de la empresa involucrada ni de la región correspondiente debido a que resultan innecesarios, ya que el ejemplo tiene fines eminentemente académicos.

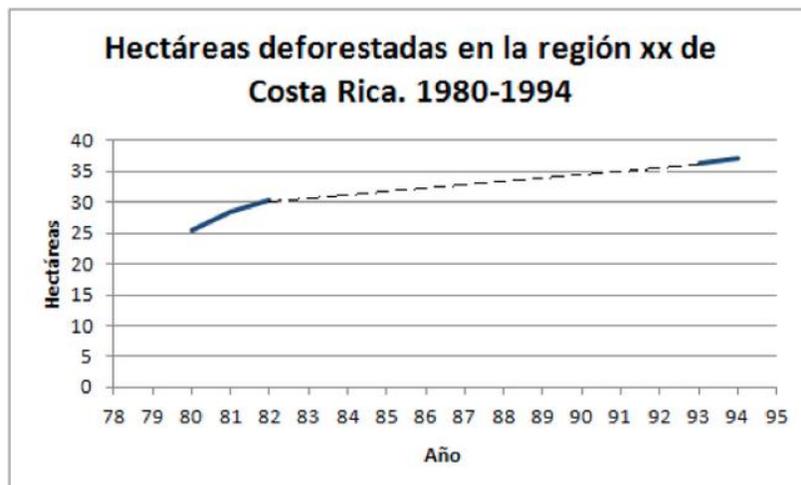
En el análisis del gráfico, la empresa consultora señala que se presentó una tasa de deforestación creciente en el período estudiado. Además resaltó una alta tasa anual de deforestación del período 1982-1993, superando las tendencias previas y posteriores a ese período.

Analice la congruencia del análisis realizado por la empresa consultora en relación con la información del gráfico y con el mensaje de los datos. ¿A qué conclusión se llega?

Análisis de actividad 2

Si se lleva a cabo una lectura rápida de la representación gráfica, pareciera que lo expuesto por la empresa es correcto. No obstante, mediante un análisis detallado se observa que el gráfico únicamente incluye información de cinco años, y con ellos se pretendió modelar el comportamiento de la deforestación para un período de quince años (1980-1994), lo que representa un abuso en el uso de la información disponible.

Además de la sobredimensión que se le estaba dando a estos cinco datos, la empresa consultora empleó en forma incorrecta la representación gráfica. Los diagramas lineales se utilizan normalmente para representar datos de series de tiempo, por lo que el tipo de gráfico es el correcto para este caso; sin embargo, cometió el error de adoptar una escala inadecuada que se inicia con unidades de un año y después utilizar la misma unidad de medida para representar un período de diez años (83-92). Por esta razón, se hace creer que la deforestación de dicho período fue más alta que en los otros años estudiados. Este intento de proporcionar información para un período en el cual no se contaba con datos constituye el principal error en que incurrieron los autores del documento. Aunque se pueda suponer que el crecimiento fue lineal en dicho período (lo cual es demasiado aventurado), la mala utilización de la escala en el eje correspondiente a los años tergiversa la información y confunde al lector. Observe ahora la representación gráfica, bajo el supuesto de linealidad y con una escala adecuada:



Fuente: Información suministrada por X.

Bajo este supuesto de linealidad, la tasa de deforestación anual en el periodo 83-92 fue mucho menor que la experimentada entre 1980 y 1982. Este ejemplo muestra que la empresa consultora pretendió extraer más información de la que los datos le podían suministrar. Hay un error adicional que cometió la empresa, al utilizar representaciones gráficas cuando la cantidad de datos es poca, lo que lleva a una generalización inadecuada.

Este ejemplo deja un importante aprendizaje: al utilizar una representación gráfica se debe cuidar la consistencia matemática de los diferentes elementos de la gráfica como las unidades de medida y escala en los ejes. Pero además, en el caso en que se cuenta con pocos datos, no se deben emplear representaciones gráficas debido a que pueden sugerir tendencias que adolecen de información suficiente para respaldarlas, lo que conlleva a generalizaciones inadecuadas.

Los ejemplos anteriores demuestran que la persona que realiza un análisis estadístico debe profundizar mucho más allá de una simple representación, debe analizar la información que está detrás de los datos en relación con el problema que le dio origen. Por esta razón, la enseñanza de la disciplina debe realizarse tomando en cuenta estos aspectos, de modo que los estudiantes puedan adquirir las habilidades necesarias para realizar esta labor.

Desafortunadamente, la enseñanza de la Estadística ha estado centrada, mayoritariamente, en la implementación de mecanismos para la descripción de hechos particulares, muchas veces descontextualizados. En la mayoría de casos, no se proponen situaciones dirigidas a la generación de un razonamiento estadístico que permita profundizar en el mensaje de la información, lo que induce a errores en el análisis de la información.

El sistema educativo nacional tiene la responsabilidad de romper con creencias equivocadas sobre diferentes fenómenos aleatorios y favorecer representaciones congruentes con los datos que se disponen; pero ante todo debe potenciar una adecuada alfabetización estadística. Esto implica promover una cultura en el manejo de la información que le permita al ciudadano enfrentar los retos de la vida cotidiana. Por ello, la acción de aula debe proporcionar a los estudiantes las habilidades necesarias para utilizar los datos del contexto con el fin de resolver problemas y tomar decisiones con argumentos sólidos y científicos.

Los actuales Programas de Estudio de Matemáticas del Ministerio de Educación Pública dejan explícita la necesidad de favorecer el proceso de enseñanza de Estadística y Probabilidad. En ellos se potencia el pensamiento aleatorio y el desarrollo de habilidades relativas para abordar situaciones de incertidumbre en la vida cotidiana, así como el uso correcto de la información para apoyar la toma de decisiones. No obstante, el reto se encuentra ahora en los salones de clase, donde se requiere implementar en la práctica de aula los fundamentos teóricos que sustentan dicha propuesta.

Seguidamente se discuten algunos aspectos de interés dentro de los procesos de alfabetización que se han venido citando.

Alfabetización en Estadística

Para llevar a cabo un proceso de alfabetización que trascienda la simple aplicación de procedimientos y técnicas, se requiere potenciar **un razonamiento estadístico** que permita al estudiante avanzar hacia la comprensión de la información que se genera en el contexto, y su vínculo con la resolución de problemas.

El profesor Carlos Araujo (s.f.), de la Pontificia Universidad Católica de Chile, resume esta cultura como la capacidad que, en materia estadística, otorga la sociedad a las personas para lograr un mejor conocimiento sobre el contexto social en el que se encuentran, así como para transmitir estas habilidades a las siguientes generaciones. La educación estadística consiste en la acción que lleva a cabo una sociedad para desarrollar, en diferentes niveles, las facultades intelectuales en el ámbito de esta disciplina. Por su parte Gal (2002), indica que se necesita alcanzar una cultura estadística que propicie al menos:

1. La capacidad para interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos estadísticos que las personas pueden encontrar en diferentes contextos.
2. La capacidad para discutir o comunicar sus opiniones respecto a tales informaciones estadísticas cuando así se requiera.

En este proceso de culturalización, los docentes responsables de llevarlo a la práctica deben cuidar tanto los aspectos disciplinares como pedagógicos relacionados con la enseñanza de la disciplina. No basta con que un estudiante pueda construir representaciones tabulares o gráficas en forma adecuada, o emplear apropiadamente los algoritmos para calcular diferentes tipos de medidas estadísticas, ya sean absolutas o relativas. Como complemento de lo anterior y en plano didáctico, se requiere propiciar habilidades de lectura, de análisis y de interpretación de los datos en concordancia con el contexto de donde provienen, sin importar la técnica estadística que se esté utilizando, de manera que no se caiga en contradicciones o ambigüedades, como ocurrió en las actividades 1 y 2. Este propósito se consigue al promover un adecuado **razonamiento estadístico** que permita trascender los aspectos rudimentales o procedimentales y lograr la resolución de los problemas con argumentos objetivos, sólidos y bien fundamentados.

Wild y Pfannkuch (1999) definieron cinco componentes básicos para lograr un adecuado razonamiento estadístico:

1. **Reconocer la necesidad de los datos:** una gran cantidad de problemas de la cotidianidad debe ser analizada considerando información pertinente, válida y confiable que otorgue evidencia concreta para decidir.
2. **Percepción de la variación:** los estudiantes deben percibir la variabilidad en los datos, la cual genera la incertidumbre sobre el mensaje que comunican. La función principal de la Estadística consiste en establecer estrategias para explicar esta variabilidad y el impacto que provoca en el análisis del problema.
3. **Transnumeración:** básicamente consiste en cambiar las representaciones de los datos para favorecer una mejor explicación de sus patrones de variabilidad. En esta etapa, se pasa de datos brutos a diferentes representaciones, en busca de aquella que pueda dar un mejor significado para el análisis que se realiza.

4. **Un conjunto de modelos:** los diferentes objetos estadísticos, sean un cuadro, un gráfico, una medida u otros más elaborados, tienen como propósito utilizar los datos para modelar su patrón de variabilidad. Se pretende lograr la sensibilidad para diferenciar los modelos de acuerdo con el tipo de dato, pero guardando las diferencias entre cada modelo y los datos mismos.
5. **Contexto, estadística y síntesis:** el razonamiento estadístico se establece al momento en que se vincula el problema generado de un contexto particular con el modelaje estadístico para realizar la síntesis de los hallazgos. Al momento de hacer estadística se puede estudiar detalladamente el comportamiento de los datos para determinar los patrones, pero estos patrones deben responder al contexto de los datos.

Estos cinco componentes son de vital importancia para la realización de análisis integrales que trasciendan la mera aplicación de técnicas. En este sentido, autores como Curcio (1989), Friel, Curcio y Bright (2001), citados por Arteaga, Batanero, Cañadas y Contreras (2011), han establecido algunos aspectos estructurales que son necesarios para llevar a cabo una adecuada interpretación de la información que, aunque se refieren a representaciones por tablas y gráficas, igualmente puede ser generalizada hacia otras técnicas. En esta literatura se menciona la relevancia que tiene para la enseñanza de la disciplina el planteo de problemas que permitan al estudiante incursionar en las siguientes etapas:

1. Leer entre los datos: consiste en llevar a cabo una lectura literal de la información, sin interpretar su contenido.
2. Leer dentro de los datos: implica no solamente interpretar los datos sino integrarlos al contexto.
3. Leer más allá de los datos: significa tomar los datos como referente para identificar patrones que trasciendan el grupo de datos observado, ya sea mediante la interpolación o extrapolación de resultados.
4. Leer detrás de los datos: consiste en llevar a cabo un análisis crítico de la información que se estudia; esto implica analizar integralmente el problema desde su origen, el tipo de dato que se utiliza, su validez y fiabilidad para analizar el problema y la posibilidad de generalizar los hallazgos.

En este sentido, la labor educativa vinculada con la enseñanza de la Estadística no solamente debe ser retadora para los estudiantes sino también para los mismos docentes, los cuales deben estar en capacidad de diseñar actividades didácticas que permitan a los jóvenes alcanzar las habilidades necesarias para lograr un adecuado razonamiento estadístico.

Existe otro elemento adicional que está inmerso dentro de la dinámica de aula y que debe ser incorporado al quehacer educativo. No hacerlo podría convertirse en una barrera didáctica que afecte los propósitos del aprendizaje. Este elemento se asocia directamente a los aspectos emotivos de los estudiantes, que son parte de ellos y fueron adquiriendo durante sus vidas, ya sea en el ámbito del hogar, la comunidad o como fruto del mismo quehacer educativo. Batanero (2002) indica que dentro de la actividad estadística intervienen diversos tipos de objetos (expresiones de lenguaje, conceptos, propiedades, acciones, argumentos) que se ponen en relación mediante correspondencias de tipo semiótico, por lo que la interpretación dada por los estudiantes no siempre concuerda con los fundamentos de la disciplina. Por ello, resulta de vital trascendencia valorar las creencias, intuiciones y actitudes que pudieran tener los estudiantes en relación con los diferentes objetos estadísticos que intervienen en el análisis de un problema. Como se observó en el ejemplo vinculado con la falacia del jugador, es significativo tomar en cuenta las creencias de las personas, de modo que, aunque

dichas creencias sean equivocadas, deben servir de motivación para generar situaciones didácticas que posibiliten modificarlas hacia un conocimiento mejor argumentado estadísticamente.

En este sentido, Fischbein (1987) establece dos tipos de intuiciones en los estudiantes que se vinculan directamente con el proceso educativo. En primer lugar, se trata de las intuiciones primarias que se adquieren con la experiencia, mediante la interacción del individuo con el contexto. Normalmente estas intuiciones no son afectadas por el proceso educativo, aunque directa o indirectamente se vinculan con conceptos teóricos que son objeto de análisis dentro de los salones de clase. Por ejemplo, la intuición sobre el azar está presente en los niños desde edades muy tempranas. No siempre estas intuiciones primarias son correctas y muchas veces se fundamentan en creencias equivocadas. En segundo lugar, Fischbein menciona las intuiciones secundarias, las cuales se forman como consecuencia del sistema educativo. Una intuición secundaria no se reduce a una fórmula simple que es aceptada y utilizada automáticamente, sino que se transforma en convicción, creencia o sentimiento de evidencia concreta sobre el concepto. Además, señala que las intuiciones secundarias no se forman a partir de la información que recibe el estudiante en un momento, ya sea por medio de una lectura o de una explicación del docente, sino que requiere de una interacción entre sus propias acciones y predicciones con el sistema educativo. A manera de ejemplo, una intuición secundaria puede ser la denominada "*falacia del jugador*", lo que responde a una mala interpretación de la "*ley de los grandes números*" que el sistema educativo no fue capaz de corregir.

En términos generales, es fundamental que los educadores tomen en cuenta las creencias e intuiciones de los estudiantes, de tal forma que el proceso educativo incida positivamente. Las actividades docentes deben transformar los aspectos negativos o erróneos y potenciar un aprendizaje significativo de los conceptos involucrados, con el fin de alcanzar habilidades para llevarlos a la práctica en forma coherente y organizada.

Seguidamente, se realiza un análisis de diferentes conocimientos estadísticos con base en los aspectos que se han venido discutiendo. Se debe observar que los problemas propuestos en el documento se han diseñado para el proceso de capacitación que se está realizando, por lo que no deben ser empleados con los estudiantes a no ser que se realice una adaptación en los casos en que sea pertinente.

Recolección de datos

En el análisis de un problema cuya solución requiere de un análisis estadístico de datos, la primera fase consiste en identificar las características de las unidades estadísticas que son objeto de estudio, así como la estrategia por medio de la cual se puede recabar información para llevar a cabo el análisis. En esta etapa, quien realiza el estudio debe verificar si alguna persona o institución ya ha recolectado información para estudiar el tema; en caso contrario deberá definir una estrategia para recolectar la información.

Actividad 3

Al inicio del Capítulo 1 del II Informe Estado de la Situación de la Persona Adulta Mayor en Costa Rica, elaborado por el Centro Centroamericano de Población (CCP), se indica:

Capítulo 1: Envejecimiento poblacional y desarrollo

Los cambios demográficos ocurridos en Costa Rica en las últimas décadas han tenido y tendrán en el futuro cercano importantes implicaciones que el país debe tomar en cuenta. El descenso del número de hijos que tienen las mujeres y el aumento en la esperanza de vida modificaron de manera importante la estructura por edad de la población. Este paso de altos a bajos niveles de mortalidad y natalidad tiene repercusiones sobre el peso relativo de los distintos grupos de edad. Al disminuir la mortalidad, un mayor número de personas sobrevive hasta edades adultas y un mayor número de personas sobrevive hasta edades avanzadas...

Fuente: http://ccp.ucr.ac.cr/espam/descargas/ESPAM_cap1web.pdf

La implicación que este hecho podría tener para el sistema educativo radica en que un descenso en la fecundidad provocaría que hacia el futuro haya menos niños en edad escolar y podría implicar una disminución en el empleo para educadores e incluso el cierre de instituciones educativas. Suponga que se requiere realizar un estudio para determinar si efectivamente se está produciendo un cambio en la fecundidad en Costa Rica.

- 1. ¿Qué tipo de información cree usted que se requiere para atender este problema?*
- 2. ¿De qué manera podría obtener información pertinente, válida y confiable para analizar el problema?*

Análisis de la actividad 3

Para realizar este análisis se necesitan de datos sobre población por edad, en algunos años anteriores al estudio, así como de proyecciones de población para los años venideros. Con esta información podría determinarse si la población en edad escolar ha venido disminuyendo con el pasar de los años y si esa tendencia se proyecta hacia el futuro.

La información necesaria para realizar estos estudios se encuentra en instituciones oficiales del país, tales como el Instituto Nacional de Estadísticas y Censos (INEC), aunque también hay instituciones no gubernamentales que pueden tener datos que ayuden a realizar el estudio. En la Universidad de Costa Rica

se encuentra el Centro Centroamericano de Población, que tiene información de calidad disponible disponible en Internet, por lo que puede ser accesada con mucha facilidad desde una computadora u otros medios electrónicos.

En este ejemplo se demuestra que no siempre los datos que se requieren para llevar a cabo un análisis estadístico deben ser recolectados por quien realiza el estudio, sino que en muchas ocasiones dicha información ya existe, y ha sido recabada por alguna institución o persona que ha estudiado el fenómeno previamente. La primera etapa de cualquier estudio consiste en determinar si existe información sobre el tema y verificar si la misma es pertinente, válida y confiable (estos conceptos serán analizados más adelante), de modo que garantice la calidad del trabajo que se va a realizar.

Actividad 4

Suponga que Karol Chaves es profesora unidocente interina en una escuela de un pequeño pueblo de zona rural. A ella le preocupa la información dada por el Centro Centroamericano de Población (que fue proporcionada en el problema anterior), en relación con el descenso de nacimientos que se ha venido presentando en el país. Esto puede provocar que su escuela sea cerrada en los próximos años y se quede sin trabajo. Si Karol desea obtener información de las familias de la comunidad donde se ubica la escuela, con el propósito de determinar la cantidad de niños menores de seis años que residen allí, de modo que pueda determinar sus expectativas de trabajo para los próximos años, ¿qué estrategia puede utilizar para seleccionar la información que requiere?

Análisis de la Actividad 4

En congruencia con lo dicho previamente, lo primero que debería hacer Karol consiste en revisar si esta información ha sido recolectada por alguna institución, por ejemplo el Centro de Salud o E.B.A.I.S. de la localidad. En caso contrario, por ser un pueblo pequeño puede realizar un sondeo entre los hogares de la comunidad, de modo que pueda conseguir la información que requiere.

Se puede observar que en los análisis de las situaciones anteriores quedan claramente diferenciadas dos estrategias para recolectar información.

1. **Información existente:** consiste en utilizar información que existe en alguna fuente conocida.
2. **Información no existente:** cuando no existe información que ha sido recolectada previamente, es necesario llevar a cabo esta labor, por medio de alguna técnica estadística que se haya diseñado con dicho propósito: observación, experimentación o interrogación (encuesta).

Esta etapa es fundamental dentro de un análisis estadístico, debido a que la calidad de los datos que se utilizan va a incidir directamente en la confiabilidad y validez de las conclusiones del estudio.

Actividad 5

Suponga que Luis Campos, director de un colegio de una zona urbano-marginal, está preocupado por el número de ausencias reiteradas de muchos de los estudiantes de su colegio, especialmente en los niveles superiores (ciclo diversificado). Tiene la creencia de que muchos de los estudiantes que se ausentan lo hacen para realizar trabajos informales y así apoyar económicamente a sus familias, por lo que estas ausencias pueden deberse a esta causa; pero además el hecho que realicen trabajos informales, aunque sea en tiempo libre, puede afectar el rendimiento académico. Por esta razón, acude a Raquel Espíndola que es profesora de séptimo año para que le ayude a investigar este hecho.

Para analizar el problema, Raquel decide aplicar un cuestionario a los estudiantes de séptimo año que se ausentaron en el último mes. Con este instrumento quiere determinar el uso que estos estudiantes dan a su tiempo libre (para establecer si algunos de ellos trabajan en ese período) y así contrastar las respuestas obtenidas. Consultó a los estudiantes que se habían ausentado al menos un día en las últimas semanas y, si la respuesta era positiva, preguntó las razones por las que se presentó este hecho.

En total 38, estudiantes se ausentaron al menos una vez en el último mes. De ellos únicamente seis señalaron que en su tiempo libre realizaban alguna actividad remunerada. Sin embargo, indicaron que sus ausencias durante el mes obedecían a problemas de salud, asistencia a citas médicas o que debieron realizar algún trámite. Estas ausencias están debidamente justificadas en el colegio. Por lo anterior, Raquel mostró al Director sus resultados, concluyendo que: la hipótesis de que los estudiantes se ausentan para trabajar pareciera ser falsa. Aun así se debe prestar atención, pues se estima que 16% de los estudiantes de séptimo año labora en su tiempo libre. Este porcentaje puede ser más alto en niveles superiores.

- a.) ¿Es correcta la conclusión a la que llegó la profesora Raquel?
- b.) En caso contrario, ¿qué errores estadísticos puede usted identificar?
- c.) Ofrezca algunas alternativas, adicionales a las que realizó la profesora Raquel, que podrían ayudar a resolver el problema con mayor propiedad.

Análisis de la actividad 5

En la solución de un problema estadístico (sin detrimento de los cuatro pasos que se deben ejecutar) se requiere identificar la unidad estadística o unidad básica del estudio, así como las características o variables que son objeto de estudio. El problema planteado por el director del colegio conlleva dos elementos básicos, por un lado se requiere identificar si algunos estudiantes del ciclo diversificado realizan trabajos informales y si, estas ocupaciones provocan que algunos de ellos deban ausentarse de lecciones. De acuerdo con este requerimiento de información, la unidad estadística es el estudiante del colegio, aunque se podría delimitar a estudiantes del ciclo diversificado para guardar relación con el planteamiento del director. Las características o variables de interés son el ausentismo y sus causas.

En esta primera etapa se detecta un primer error cometido por la profesora, debido a que cambió la unidad estadística y realizó el estudio con los estudiantes de séptimo año lo que provoca que los resultados no sean *pertinentes* en relación con el problema original. Por las diferencias de edades entre estudiantes de séptimo año con respecto a los que están en décimo o undécimo, es de esperar que existan también diferencias en las actividades que realizan en su tiempo libre.

En segundo lugar, Raquel decide delimitar el tiempo de estudio, lo cual es adecuado para el análisis de este tipo de problemas, pero podría ser que un mes no sea el tiempo apropiado para llevar a cabo esta investigación.

Por otro lado, la estrategia de recolección de información empleada por la profesora consistió en identificar las actividades que los estudiantes realizan en su tiempo libre. Dicha estrategia se fundamenta en el supuesto de que los estudiantes que realizan actividades laborales en este tiempo son más propensos a ausentarse de lecciones para trabajar. Esto puede acarrear un problema de *validez* de la información, pues el hecho de que los estudiantes trabajen en su tiempo libre no necesariamente responde a la interrogante del problema: *¿se ausentan los estudiantes para realizar trabajos informales?* Aunque consultó si el motivo de las ausencias obedecía a razones laborales, nuevamente se debe cuestionar la validez de las respuestas, debido a que el ausentarse del colegio para trabajar genera ausencias injustificadas, mientras que hacerlo por razones médicas es justificable. Esta amenaza conlleva que los estudiantes pudieran estar mintiendo en sus respuestas.

Por otro lado, la selección de los estudiantes que se habían ausentado en el último mes no es adecuada para concluir que *el 16% de todos los estudiantes de séptimo año labora en su tiempo libre*. Para llegar a una aproximación o estimación de este tipo, se requiere de una muestra aleatoria que sea representativa de todos los estudiantes del nivel. La generalización realizada por la profesora Raquel lleva implícito un abuso de la información. Con los datos generados desde un pequeño grupo de estudiantes que no son representativos de todos los séptimos años, no es prudente realizar una estimación hacia todos los estudiantes del nivel. Por esta razón, la conclusión no es correcta debido a que los datos no son *confiables* para llegar a ella.

Datos pertinentes, válidos y confiables

La solución de la actividad 5 incluyó en la discusión una serie de elementos que resultan de gran relevancia dentro de los análisis estadísticos. Se han citado tres términos vinculados con la recolección de datos: ***pertinencia, validez y confiabilidad***.

La pertinencia

El primer error que cometió la profesora Raquel consistió en utilizar información de alumnos de séptimo año para analizar un problema que estaba planteado para estudiantes del ciclo diversificado. En este caso se dice que los datos empleados no eran pertinentes para el problema en estudio.

En general, se dice que los datos que se utilizan para un estudio son ***pertinentes*** si responden a los intereses del problema, de manera que con ellos se puede contar con información que sea congruente con las interrogantes que genera dicho problema.

La validez

Los datos utilizados en el problema para identificar si los estudiantes se ausentaban a lecciones para realizar trabajos informales resultaron no ser válidos debido a que respondían a una interrogante diferente a la original. La profesora Raquel consultó a los estudiantes que se ausentaron en el último mes si trabajaban en su tiempo libre. Las respuestas a estas interrogantes no responden necesariamente el hecho de que se ausenten a lecciones para trabajar.

Otro problema de *validez* que presenta el estudio realizado por la profesora Raquel, se relacionó con el hecho de que para el análisis de temas delicados como el actual, no es conveniente realizar la pregunta en forma directa. Es decir, en este caso no se les debería preguntar a los estudiantes si se han ausentado de clases para realizar actividades laborales pues obviamente la respuesta va a ser negativa. Las respuestas obtenidas no necesariamente responden a los requerimientos de información.

En general, la validez de los datos hace referencia al grado en que ellos miden la característica o variable que se pretende medir. Esto no siempre ocurre; por ejemplo, cuando se aplica un test de inteligencia, los resultados dejan de ser válidos si los datos en vez de medir inteligencia miden la memoria u otra característica diferente.

Hay variables para las cuales es muy fácil determinar datos válidos, por ejemplo: nivel educativo, lugar de residencia, entre otras. Estas son preguntas concretas directas a las unidades estadísticas, que normalmente generan datos válidos. Pero existen variables, vinculadas con emociones, creencias o actitudes, en que se requiere realizar pruebas estadísticas de validez, para garantizar que la información recabada corresponda con aquello que se quiere medir.

La confiabilidad

Para que el estudio realizado por Raquel hubiera generado datos confiables, debía aplicar un procedimiento de muestreo aleatorio que le garantice una muestra representativa de la totalidad de estudiantes de séptimo año, en el sentido de que si repitiera el estudio posteriormente y seleccionara otra muestra, los resultados deberían ser similares. Pero esto no se presentó en el caso del análisis realizado, pues la muestra escogida ni es aleatoria ni es representativa. Por ello, la conclusión de que el *16% de los estudiantes de séptimo año labora en su tiempo libre* no es confiable.

En general, la confiabilidad de una técnica de recolección de información hace referencia al nivel de precisión o exactitud con que se recolectan los datos. Es decir, se espera que al aplicar la técnica dos o más veces a un mismo fenómeno se obtengan resultados idénticos. Por ejemplo, si una persona se realiza una prueba de sangre para determinar el nivel de colesterol, los resultados son confiables si en el laboratorio realizan dos o más veces el análisis y obtienen resultados similares. En el caso de la aplicación de un cuestionario, se dice que es confiable si al implementarlo dos veces a un determinado sector de la población los resultados generados son similares.

Análisis didáctico

Existen diversas habilidades específicas en los Programas de Estudio de Matemáticas para secundaria (páginas 354-356), vinculadas con los aspectos discutidos en esta sección. Por ejemplo,

Sétimo año	Habilidad
Conocimientos básicos <ul style="list-style-type: none"> · Unidad estadística · Características · Datos u observaciones · Población · Muestra 	Identificar los conceptos: unidad estadística, características o variables, observaciones o datos, población y muestra, para problemas estadísticos vinculados con diferentes contextos.
Recolección de información <ul style="list-style-type: none"> · La experimentación · Interrogación 	Recolectar datos del entorno por medio de experimentación o interrogación.
Octavo año	Habilidad
Recolección de información <ul style="list-style-type: none"> · La experimentación · Interrogación 	Recolectar datos del entorno por medio de experimentación o interrogación.

El docente debería proponer problemas que permitan a los estudiantes vincular algunos de los elementos discutidos en las actividades de esta sección con el fin de favorecer las habilidades descritas. Se debe tener claro que dichas actividades fueron diseñadas para los docentes participantes en la capacitación y no para desarrollar con los estudiantes. Por esta razón, si se desea implementar alguna de ellas se requeriría llevar a cabo modificaciones pertinentes. Por ejemplo, la actividad 5 puede ser modificada para trabajar con estudiantes de sétimo año, de la siguiente manera:

Variación de la Actividad 5

Suponga que don Luis Campos es el director de un colegio de una zona urbano-marginal, y está preocupado por el número de ausencias reiteradas de muchos de los estudiantes de su colegio, especialmente en los niveles superiores (ciclo diversificado). Tiene la creencia de que muchos de los estudiantes de estos niveles realizan trabajos informales para apoyar económicamente a sus familias, por lo que estas ausencias pueden deberse a esta causa; pero además el hecho que realicen trabajos informales, aunque sea en tiempo libre, puede afectar el rendimiento académico. Por esta razón, acude a una de las profesoras de Matemáticas (llamada Raquel) para que analice el problema. Ella es profesora guía de la sección 7-3

Para atender esta solicitud, la profesora decidió aplicar un cuestionario a los estudiantes que se ausentaron a lecciones en el último mes en su sección guía. Con este cuestionario quiere determinar el uso que estos estudiantes dan a su tiempo libre (para determinar si algunos de ellos trabajan en ese tiempo). Entre las preguntas que realiza están las siguientes:

1. Indique el número de horas por día que, aproximadamente, dedica a cada una de las siguientes actividades:

Actividades	Número de horas
Ver televisión	
Uso de computadora para Internet	
Escuchar música	
Juegos de videos	
Hacer tareas escolares	
Realizar trabajos en el hogar	
Realizar trabajos fuera del hogar	
Otras	

2. ¿Ha tenido que ausentarse del colegio al menos un día en el último mes?

() Si () No (pase a la pregunta 4)

3. En caso que se haya ausentado a lecciones, indique las razones:

4. (...)

De los 38 estudiantes de la sección 7-3 se ausentaron 12 al menos una vez en el último mes. A estos 12 estudiantes les aplicó el cuestionario, de los cuales seis dijeron que en su tiempo libre realizaban alguna actividad que podría ser catalogada como laboral. Sin embargo, indicaron que sus ausencias durante el mes obedecían a problemas de salud, que debían asistir a alguna cita médica o que debieron realizar algún otro trámite. Estas ausencias están debidamente justificadas en el colegio. Por lo anterior, la profesora Raquel mostró al Director sus resultados y concluyó que la idea de que los estudiantes se ausentan para trabajar pareciera ser falsa. Aun así se debe prestar atención, pues encontró que seis de los 38 estudiantes (aproximadamente 16%) de su sección laboran en su tiempo libre, lo que puede afectar su rendimiento en el colegio.

1. *¿Son correctas las conclusiones a las que llegó la profesora Raquel,*
 - a. *de que no hay estudiantes de séptimo año que se ausenten para laborar?*
 - b. *de que aproximadamente el 16% de los estudiantes de la sección 7-3 labora en su tiempo libre?*
2. *Si no fueran correctas estas conclusiones, ¿qué errores cree que pudo haber cometido esta profesora?*
3. *Ofrezca algunas alternativas, adicionales a las que realizó la profesora Raquel, que podrían analizar de una mejor manera el problema.*

Mediante esta actividad se potencian las siguientes habilidades específicas de séptimo año:

- Identificar los conceptos: unidad estadística, características o variables, observaciones o datos, población y muestra, para problemas estadísticos vinculados con diferentes contextos.
- Recolectar datos del entorno por medio de experimentación o interrogación.

El problema planteado permite integrar diferentes procesos matemáticos y ejes disciplinares. Los estudiantes deben reflexionar sobre la estrategia empleada por la profesora y encontrar posibles errores en el procedimiento. El razonamiento y la argumentación son componentes básicos que deben estar presentes durante la etapa de análisis.

En la etapa de clausura o cierre, el docente debe repasar las soluciones dadas por los estudiantes, con el propósito de vincular el trabajo realizado en el problema con los conocimientos y habilidades que deben ser debatidos. Al menos deberían ser considerados los siguientes elementos:

1. *Unidad estadística*: se debe identificar que la unidad estadística corresponde al estudiante de séptimo año del colegio. Hacer hincapié en la definición general de unidad estadística en cualquier problema. A diferencia del problema planteado en la actividad 5, en esta actividad la unidad estadística ha sido bien concebida, pues corresponde a los estudiantes de séptimo año.
2. *Características o variables*: es importante identificar qué variables utilizó la profesora Raquel para el análisis efectuado. Entre otras se pueden citar todas aquellas que se vinculan con las actividades del estudiante en su tiempo libre: el tiempo que dedica a cada actividad, ausentismo (sí o no), causas para ausentarse. Como Raquel solamente decidió aplicar el cuestionario a los estudiantes que se han ausentado, entonces se espera que, en todos los casos, la respuesta al ausentismo sea positiva. Para comprender mejor estos conceptos, se puede recurrir a la forma en que se resumiría esta información: una base de datos con el siguiente esquema.

Estudiante	Ver TV	Internet	Música	Juegos	Deporte	Trabajos escolares	Trabajos hogar	Trabajos fuera	Otras	Ausencia	Causas
1	3	1	2	0	0	1	0	0	0	Si	Enfermedad
2	1	0	0	0	0	2	0	4	0	Si	Cita medica
⋮											

Las unidades estadísticas son los estudiantes, cada uno debe proporcionar información de 11 variables. La información proporcionada por estos corresponde a los datos.

3. *Población*: este concepto es fundamental para el análisis del problema, debido a que el error cometido por la profesora Raquel obedece a que la población es la totalidad de los estudiantes de séptimo año.
4. *Muestra*: la muestra que utilizó la profesora Raquel no es representativa de la población, ella decidió escoger su grupo guía, por lo que es una muestra seleccionada por conveniencia. Si se quería analizar las razones del ausentismo entre los estudiantes de séptimo año, se debía escoger una muestra representativa de todas las secciones y no simplemente de un grupo, como lo hizo ella.

Muestreo estadístico e inferencia estadística

Al igual que la técnica de recolección de información, la identificación de las unidades estadísticas que van a ser seleccionadas para realizar el estudio constituye un pilar fundamental de los procesos de análisis estadístico. En la actividad 5 se evidenciaron problemas para seleccionar a los estudiantes en la estrategia utilizada por la profesora Raquel. Observe el siguiente problema:

Actividad 6

El Ministerio de Educación Pública le pide a la Dirección de un colegio seleccionar una muestra de 10 estudiantes de noveno año para que participen en una prueba internacional de Matemáticas. La Dirección asigna esta selección, de los diez estudiantes, a los profesores de noveno año: Marcelina Solís, Guillermo Brenes y Mynor Mora., para que sean ellos los que realicen la selección. Pero advierte que la selección realizada debe representar adecuadamente a la totalidad de estudiantes de noveno año del colegio. La institución cuenta con 10 grupos de noveno año, todos ellos con diferente cantidad de estudiantes. Estos tres profesores se reúnen y cada uno de ellos hace una propuesta distinta para realizar esta selección:

- 1. La profesora Marcelina indica que para garantizar una adecuada representación por sexo, deberían seleccionar los cinco varones y cinco mujeres que han tenido los mayores rendimientos en Matemáticas durante el año.*
- 2. El profesor Guillermo señala que al existir 10 grupos lo mejor es seleccionar un estudiante de cada grupo, para garantizar aleatoriedad. Propone escoger un número fijo (por ejemplo el número 15) y se selecciona el estudiante que tiene este número asignado en cada lista. El nombre de este estudiante sería desconocido, debido a que no se tienen las listas en este momento.*
- 3. Finalmente, el profesor Mynor indica que le gusta la idea de Guillermo, pero que lo ideal sería que el estudiante de cada grupo sea seleccionado mediante una rifa entre los estudiantes que tienen un rendimiento cercano al promedio de cada grupo. Con esto se evita que sean escogidos los estudiantes con peores o mejores rendimientos, lo que garantizaría objetividad.*

Analice las propuestas. ¿Cuál de ellas genera una muestra más representativa de la totalidad de estudiantes de noveno año? ¿Podría usted sugerir alguna estrategia que garantice una mayor representatividad?

Análisis de la actividad 6

Los tres profesores propusieron una estrategia sencilla y fácil de llevar a la práctica; sin embargo, para decidir cuál de ellas es más adecuada para resolver el problema es necesario revisar la objetividad y aleatoriedad de cada una.

En primer lugar, la profesora Marcelina propone garantizar la representación por sexo, lo cual no es problema. No obstante, sugiere también que se prefieran los estudiantes que han tenido los mejores resultados durante el año escolar. Esta selección no es representativa de la totalidad de estudiantes, debido a que solamente aquellos que han tenido excelente rendimiento podrían ser elegidos para participar en la prueba internacional.

Por su parte, el profesor Guillermo considera seleccionar un estudiante por grupo, lo que ofrece representación de todos los grupos. Para garantizar la aleatoriedad propone que se escoja el estudiante que tiene una posición fija en la lista en cada uno de los grupos. En primera instancia, esta propuesta parece adecuada, pues al escoger un número fijo de la lista, por ejemplo el número 15, se escogerían 10 estudiantes que ocupan esta posición. Como el número es indeterminado, pareciera que la selección podría ser aleatoria. Pese a ello, no indica el profesor Guillermo la forma en que se va a escoger este número: si se realiza por criterios subjetivos se perdería la aleatoriedad. Pero además, como los grupos son de diferente tamaño, el número escogido no puede superar al menor número de estudiantes por grupo, con lo cual los estudiantes que ocupan en la lista de clase un número superior a este valor mínimo no tendrían oportunidad de ser seleccionados.

En relación con el planteamiento del profesor Mynor, esta la propuesta del profesor Guillermo (escoger un estudiante por grupo), pero varía la forma de selección del estudiante. Al proponer una selección por medio de una rifa, consigue la aleatoriedad que no tenía la propuesta de Guillermo, pero decide eliminar arbitrariamente a los estudiantes que tienen calificaciones muy por debajo o muy por encima de la calificación promedio del grupo. Nuevamente la selección no es representativa de los todos los estudiantes de noveno año.

Como puede notarse, todas las propuestas presentan algún problema de representatividad. Si tuviera que escogerse alguna de ellas, posiblemente la que tiene menos problemas de representatividad es la que realizó el profesor Guillermo Brenes.

Sin embargo, para realizar una selección aleatoria deben considerarse otras posibilidades, entre las que se pueden citar:

1. Tomar en cuenta la propuesta del profesor Brenes de seleccionar un estudiante por sección y la propuesta del profesor Solís de realizar una rifa para llevar a cabo la escogencia pero incluyendo a todos los estudiantes de cada grupo. Esta forma de realizar el muestreo garantiza que todos los estudiantes tienen una probabilidad positiva de salir seleccionados.
2. La propuesta anterior asigna probabilidades de salir seleccionados a los estudiantes, pero no es la misma para todos, debido a que al ser grupos de diferente tamaño es más probable salir electo en los grupos más pequeños y menos probable en los grupos más grandes. Si se desea que todos los estudiantes tengan la misma probabilidad de salir seleccionados, la forma más simple consiste en elaborar un listado con la totalidad de estudiantes de noveno año y seleccionar aleatoriamente a 10 de ellos, lo cual se podría realizar mediante una rifa.

El problema anterior plantea un ejemplo de muestreo estadístico simple. En general, las técnicas de muestreo estadístico pueden ser complejas, por lo que muchos problemas requieren de la participación de especialistas en el campo que puedan elaborar un diseño muestral apropiado. En este documento se discuten algunas herramientas elementales sobre el tema.

Tipos de muestreo

Tradicionalmente, las técnicas de muestreo estadístico se clasifican en dos grandes grupos: técnicas de **muestreo probabilístico o aleatorio** y técnicas de **muestreo no probabilístico**.

Muestreo probabilístico o aleatorio

Las técnicas de muestreo probabilístico se basan en procesos de selección por medio de los cuales todas las unidades estadísticas de la población tienen asignada una probabilidad positiva de ser escogidas en la muestra. Como puede notarse, las dos técnicas propuestas al final del análisis de la actividad 6 constituyen técnicas de muestreo probabilístico. Las técnicas de muestreo aleatorio garantizan la representatividad de la población y, mediante técnicas de inferencia estadística, sus resultados pueden utilizarse para ser generalizados hacia la población. Existen diferentes técnicas de muestreo probabilístico:

- 1. Muestreo simple al azar:** consiste en la técnica más simple de muestreo aleatorio. Todas las unidades estadísticas de la población tienen la misma probabilidad de ser escogidas. Normalmente se realiza cuando se tiene un listado o enumeración de todas las unidades estadísticas de la población y mediante una selección aleatoria (por ejemplo una rifa) se escogen “n” elementos que constituyen la muestra.

A manera de ejemplo, suponga que en el problema anterior se decide seleccionar a los diez estudiantes escogiéndolos en forma aleatoria del listado completo de noveno año. Si en total hubiera 312 estudiantes, la población de estudiantes estaría representada en un listado que va enumerado del número 1 al 312. La aleatoriedad se presenta al momento de que la selección se realice por medio de alguna estrategia que garantice que todos tengan la misma probabilidad de ser escogidos, la cual puede ser mediante una rifa u otro procedimiento similar. En este caso la probabilidad de selección para cada estudiante sería $\frac{10}{312}$.

- 2. Muestreo sistemático:** esta técnica, al igual que la anterior, requiere enumerar todos los “N” elementos o unidades estadísticas de la población, pero en lugar de seleccionar aleatoriamente los “n” elementos, solamente se extrae aleatoriamente uno de los primeros k elementos. Los restantes se seleccionan a partir de este en forma sistemática mediante un ciclo de k elementos. El valor de k está dado por $k = \frac{N}{n}$.

Como ejemplo, suponga de nuevo que se deben seleccionar 10 del total de los 312 estudiantes de noveno año. En este caso $k = \frac{312}{10} = 31,2$. Entonces el primer estudiante se selecciona al escoger un número entero aleatorio entre 1 y 31. Suponga que el número aleatorio es M, entonces la selección de los estudiantes corresponde a la sucesión de números $M + k \cdot i$, donde $i \in \{1, 2, \dots, 10\}$. Por ejemplo, si el número aleatorio entre 1 y 31 es el 17, entonces la sucesión correspondiente sería:

17,0 48,2 79,4 110,6 141,8 173,0 204,2 235,4 266,6 297,8

Los números se redondean a la unidad inmediata (superior o inferior según corresponda). Esto proporciona las posiciones de los estudiantes que se deben seleccionar. A pesar de la mayor complejidad de esta técnica, se preserva la probabilidad de selección para cada estudiante de $\frac{10}{312}$.

Esta técnica tiene la virtud que, como su nombre lo indica, en la escogencia se realiza un recorrido sistemático por todo el listado. Al utilizar un muestreo simple al azar de pocas unidades estadísticas

(como en el presente ejemplo) puede ocurrir que, por azar, haya una mayor selección de números bajos, altos o intermedios; con el muestreo sistemático se evita este problema de concentración aleatoria. En el contexto del problema garantizaría (si la lista general de estudiantes va en forma secuencial por grupo), que la selección evitaría la concentración aleatoria de estudiantes de un mismo grupo o grupos.

Un análisis más detallado de la técnica sobrepasa los propósitos de este texto.

3. Muestreo estratificado: esta es una técnica más compleja. Se utiliza cuando la población está congregada en forma natural en pocos grupos o estratos. Entonces se selecciona una muestra aleatoria en cada estrato. La muestra general está constituida por las muestras particulares. Este tipo de muestreo reduce los errores de estimación, pero requiere realizar ponderaciones para los diferentes cálculos que se vayan a efectuar.

Suponga por ejemplo que el Circuito Educativo 03 de una Dirección Regional de Costa Rica tiene tres colegios de dimensiones diferentes. El Colegio A tiene 546 estudiantes, el Colegio B tiene 933 y el Colegio C tiene 1542. Se solicita a las autoridades de la Dirección Regional llevar a cabo una selección aleatoria de 50 estudiantes de secundaria para que completen una encuesta. Aunque se pueden utilizar las técnicas anteriores para la escogencia de esta muestra, el muestreo estratificado simplifica el proceso. Debido a que las tres instituciones son de diferente tamaño, lo ideal sería seleccionar un número de estudiantes en cada colegio en forma proporcional al tamaño de la institución (aunque no es la única forma). En el siguiente cuadro se indica el número de estudiantes que se debería seleccionar en cada colegio:

Institución	Número de estudiantes	Porcentaje "p"	Proporción de muestra $\frac{p}{100} \cdot 50$	Muestra
A	546	18,1	9,05	9
B	933	30,9	15,45	15
C	1542	51,0	25,50	26
Total	3021	100,0	50	50

Entonces deben seleccionarse 9 estudiantes del Colegio A, 15 del B y 26 del C. La selección en cada institución puede realizarse en forma simple al azar o en forma sistemática.

Debe recalcar que al emplear un muestreo estratificado no basta con juntar las muestras de cada estrato y generar una muestra mayor.

La complejidad de aplicar esta técnica no se centra en la determinación de las muestras que deben seleccionarse en cada estrato, lo cual es relativamente simple tal y como se observó en este ejemplo. Para aprovechar estadísticamente la estratificación, se requiere realizar un análisis estadístico en cada estrato y luego juntar los resultados para determinar el comportamiento de la muestra total. Sin embargo, en este proceso debe ponderarse el aporte de cada estrato tanto por su tamaño como por la variabilidad que cada uno contiene.

Un análisis más detallado sobre el tema sobrepasa los propósitos del curso. Si algún lector desea profundizar se recomienda consultar algún texto sobre muestreo estadístico; dos de los mejores textos en el campo son:

Cochran, W. G. *Sampling Techniques*, 3rd Edn., Wiley, New York. 1977

Kish, L. *Muestreo de encuestas*. Traducido al español por Ricardo Vinós Cruz López. Editorial Trillas. México, D.F., México. 1975.

4. **Muestreo por conglomerados:** a diferencia de las técnicas anteriores, en el muestreo por conglomerados no se seleccionan directamente las unidades estadísticas, sino que se seleccionan grupos o conglomerados de unidades estadísticas. Una vez seleccionado un conglomerado se requiere seleccionar las unidades estadísticas dentro de él (para lo cual puede utilizar muestreo simple al azar o muestreo sistemático). Esta técnica es mucho más compleja que las anteriores, pero es la más utilizada en estudios estadísticos sofisticados que involucran poblaciones grandes y complejas, como por ejemplo las encuestas nacionales. El tipo de conglomerado más utilizado en la investigación es el geográfico: consiste en dividir la población en zonas geográficas. Aunque dependiendo del caso, se pueden utilizar otros criterios.

Por ejemplo, un investigador desea estudiar el rendimiento académico de los estudiantes de secundaria en Costa Rica. Entonces se puede proceder de la siguiente manera:

- Puede dividir a toda la población (estudiantes de secundaria de todo el país) en diferentes conglomerados que podrían ser los circuitos educativos.
- Luego, en forma aleatoria selecciona una cantidad adecuada (previamente fijada) de circuitos, para ello puede combinar el muestreo aleatorio y los muestreos simple al azar o sistemático.
- Una vez escogidos los circuitos educativos (conglomerados) debe realizar una selección de estudiantes en cada circuito, para lo cual puede nuevamente combinar el muestreo estratificado y los muestreos simple al azar o sistemático. Si el número de instituciones en algunos circuitos es muy grande, puede nuevamente aplicar muestreo por conglomerados dentro de los circuitos educativos.

Lo más relevante sobre esta técnica de muestreo consiste en que todos los conglomerados tengan una probabilidad positiva de ser seleccionados.

Al igual que en la técnica anterior, no basta con juntar todas las muestras y generar una muestra general, sino que debe trabajarse a lo interno de cada conglomerado y mediante técnicas de ponderación. Los resultados deben conjuntarse para el análisis integral de la muestra.

Por las mismas razones señaladas que se indicó en el caso del muestreo estratificado, un análisis detallado del muestreo por conglomerados sobrepasa los propósitos del curso. Si el lector está interesado en el tema puede consultar la literatura citada.

Existen otras variantes de las técnicas anteriores vinculadas con el muestreo aleatorio; pese a ello, se ha considerado conveniente no incluirlas en la discusión, debido a que las estrategias discutidas antes son suficientemente representativas.

Muestreo no probabilístico

En las técnicas de muestreo no probabilístico los elementos de la población no tienen asignada una probabilidad positiva de selección, o sea no todos ellos tienen posibilidad de ser seleccionados en la muestra. Por esta razón, dichas técnicas no garantizan la representatividad de la población, y no deben utilizarse para generalizar los resultados; se utilizan simplemente para realizar análisis descriptivos entre los elementos de la muestra. Existen muchas técnicas de muestreo no probabilístico, las más comunes son:

1. **Muestreo por conveniencia:** probablemente el muestreo por conveniencia sea la técnica de muestreo que mayoritariamente se utiliza. Las unidades estadísticas son seleccionadas porque son accesibles para la persona que realiza el estudio. Es decir, los elementos se eligen porque son de fácil ubicación, lo que simplifica el análisis.

Por ejemplo, Martín es un profesor de Matemáticas que debe realizar un estudio para determinar la opinión de los estudiantes del colegio sobre el uso de la calculadora en las clases de Matemáticas. Diseña un cuestionario y lo aplica a los estudiantes de su grupo guía de noveno año. Los resultados del estudio solamente van a ser válidos para este grupo.

Normalmente, el muestreo por conveniencia genera muestras que son poco representativas de la población general.

2. **Muestreo por juicio de expertos o discrecional:** consiste en tomar en cuenta el criterio de expertos en el área de estudio; son ellos los que deciden o recomiendan qué unidades estadísticas deberían ser incluidas en la muestra, de modo que estén en capacidad de dar un aporte sustancial al estudio.

En el ejemplo planteado en el punto anterior, Martín puede pedirles a sus colegas que seleccionen dos o tres estudiantes de cada grupo que tengan un criterio claro sobre el significado de realizar un uso inteligente de la calculadora en las clases de Matemáticas, labor que también Martín realiza en sus grupos. Al final, se obtiene una muestra de estudiantes cuyos aportes pueden ser de mucho interés para el estudio.

Ninguno de estas dos técnicas genera muestras representativas de la población estudiantil, como puede ocurrir si se aplica un muestreo aleatorio

3. **Muestreo por cuotas:** consiste en la selección de elementos de la población que tienen ciertas características en busca de una buena representación. La selección de los elementos o unidades estadísticas se realiza por conveniencia (no aleatoria), pero se busca que los elementos seleccionados formen parte de una cuota.

Por ejemplo, en el caso de la actividad 6, una forma de aplicar esta técnica podría ser considerando tres agrupaciones de estudiantes: estudiantes de bajo rendimiento, estudiantes de rendimiento aceptable (medio) y estudiantes de alto rendimiento. Luego de llevar a cabo la clasificación, se debe seleccionar una cuota de estudiantes de cada grupo en forma proporcional a la cantidad de estudiantes que se encuentra en cada condición de rendimiento. Se supone que en total habían 297 estudiantes de noveno año, de los cuales 57 (19,2%) tenían bajo rendimiento, 158 (53,2%) tenían rendimiento promedio (aceptable) y 82 (27,6%) un rendimiento alto. La selección que debería realizarse se incluye en el siguiente cuadro:

<i>Rendimiento</i>	<i>Población</i>	<i>Porcentaje</i>	<i>Muestra $n = 10$ (cuotas)</i>
Bajo	57	19,2	$0,192 \cdot 10 = 1,92 \approx 2$
Medio	158	53,2	$0,532 \cdot 10 = 5,32 \approx 5$
Alto	82	27,6	$0,276 \cdot 10 = 2,76 \approx 3$
Total	297	100,0	10

Por lo tanto las cuotas correspondientes son: dos estudiantes de bajo rendimiento, cinco estudiantes de rendimiento medio y tres estudiantes de alto rendimiento. La selección de la cuota en cada grupo se realiza en forma no aleatoria, ya sea por conveniencia o a juicio de los profesores.

Observe que esta técnica puede utilizarse en forma parecida al muestreo por conglomerados, solo que la selección de las unidades estadísticas no es aleatoria. Si en este ejemplo la selección de los estudiantes de cada grupo se hubiera realizado en forma aleatoria, por ejemplo, mediante muestreo simple al azar, entonces el muestreo aplicado sería aleatorio por conglomerados.

Existen otras alternativas de selección de muestras no probabilísticas, pero las que se han planteado acá resultan ser la más utilizadas.

La importancia de utilizar muestras aleatorias, además de generar una mejor representación de la población, obedece a que solamente mediante el uso de muestras aleatorias es posible inferir hacia la población. Esto quiere decir que los resultados generados en una muestra aleatoria de unidades estadísticas pueden ser generalizados hacia la población, sujeto a márgenes de **error aleatorio** que pueden ser medidos probabilísticamente, tal como se discutirá más adelante. Por medio de estas muestras, pueden estimarse valores poblacionales o incluso probar hipótesis sobre ellos.

Estas inferencias hacia la población no se pueden realizar con muestras no aleatorias, debido a que los procesos de generalización de resultados se fundamentan en la teoría de probabilidades; pero en el caso de estas muestras la selección de las unidades estadísticas no se realiza bajo principios probabilísticos, tal como se describió arriba. Por esta razón, las conclusiones de los análisis con muestras no probabilísticas únicamente son válidas para el grupo de unidades estadísticas con las que se está trabajando, es decir

corresponde a estudios meramente descriptivos de este grupo.

Con el propósito de aclarar un poco más estos aspectos a aquellos lectores que no están familiarizados con el tema, seguidamente se incluye una descripción de los estudios estadísticos descriptivos y los inferenciales.

Estadística descriptiva y estadística inferencial

Como se ha venido indicando, la Estadística es una disciplina que se ha desarrollado para ayudar a comprender el patrón de variabilidad de los datos que se generan en otras áreas científicas. Para realizar esta labor se utilizan diferentes técnicas que involucran procesos de recolección, organización, representación, determinación de medidas sobre un conjunto de datos con el propósito de caracterizar su patrón de variabilidad y extraer conclusiones que sean adecuadas para un problema concreto. Muchas veces, los análisis estadísticos mediante el empleo de procesos probabilísticos permiten realizar generalizaciones hacia la población de interés, sujetos a ciertos márgenes de error. En este sentido, la Estadística se puede clasificar en Estadística Descriptiva o Estadística Inferencial.

1. **Estadística descriptiva:** es una parte de la Estadística que se dedica a organizar, representar, resumir y analizar conjuntos de datos, de modo que se puedan describir en forma precisa sus patrones de variabilidad e identificar las principales características en función del contexto al que pertenecen. Este análisis es fundamental en cualquier estudio que requiera de datos para la toma de decisiones.

En esta etapa, la confección de cuadros y gráficas, el cálculo de medidas de posición como de variabilidad y el cálculo de porcentajes son algunas de las técnicas comúnmente utilizadas para favorecer la interpretación, el análisis y la argumentación que se genera de los datos. Cabe rescatar que los distintos tipos de representación de datos son un apoyo para la interpretación y el análisis de los mismos; sus elaboraciones o procedimientos de cálculo no son el fin sino el medio para realizar los análisis estadísticos.

2. **Estadística inferencial:** la información proporcionada por los análisis descriptivos en una muestra aleatoria de datos, puede ser aprovechada para extraer conclusiones sobre la población de donde se extrajo esa muestra. Esta capacidad para poder inferir sobre la población con base en los resultados de una muestra aleatoria está basada en supuestos probabilísticos relacionados con los patrones de variabilidad presentes en la distribución de los datos. El conjunto de técnicas y procedimientos por medio de los cuales se vinculan los resultados de la estadística descriptiva con la aplicación de modelos de probabilidad para inferir o generalizar esos resultados es denominado Inferencia estadística o Estadística inferencial.

Es significativo hacer notar que la Estadística y la Probabilidad surgen a través de la historia como dos ramas completamente separadas; años después se descubre que si se unen pueden crear áreas nuevas de conocimiento. La Estadística inferencial es el vínculo entre la Estadística y la Probabilidad, aquí se trabaja con muestras obtenidas bajo procesos probabilísticos y gracias a esta obtención se pueden hacer generalizaciones a las poblaciones en estudio, permitiendo identificar la magnitud máxima de los errores aleatorios, encontrar intervalos en los que pueden encontrarse los valores poblacionales e incluso probar

supuestos que han sido preestablecidos sobre estos valores poblacionales.

Las técnicas empleadas para realizar los análisis inferenciales sobrepasan los propósitos del curso, pero es fundamental que los docentes de Matemáticas conozcan estas diferencias entre lo descriptivo y lo inferencial, debido a que son objeto de discusión dentro de la actividad educativa que realizan.

El término **error** se ha venido aplicando en diferentes partes del documento. En esta última parte se ha indicado que mediante muestras aleatorias es posible determinar el *error aleatorio o error de muestreo*. Por esta razón resulta de interés para el lector comprender su significado, aunque los procedimientos de cálculo no se incluyan acá. Resuelva el siguiente problema.

Actividad 7

Estudiantes de la Escuela de Estadística de la Universidad de Costa Rica realizaron un estudio a través del cual, entre otras cosas estimaron el porcentaje de seguidores de los equipos de fútbol de la primera división del país. El periódico La Nación publicó la siguiente información:

En Costa Rica hay igual cantidad de liguistas y de saprissistas

POR CAROLINA RUIZ V. - Actualizado el 29 de enero de 2013 a: 12:00 a.m.

San José (Redacción). Empate técnico. En el país hay igual cantidad de liguistas que de saprissistas, reveló una encuesta de la Universidad de Costa Rica (UCR).

Luego de consultar a 842 personas, como parte de la Encuesta Actualidades 2012, dos estudiantes de la Escuela de Estadística de la UCR descubrieron que 35,9% son aficionados a la Liga Deportiva Alajuelense (LDA) y 33,3%, al Deportivo Saprissa.

Al tomar en cuenta el error de muestreo de la estimación, se estima el empate técnico entre ambas barras de seguidores.



http://www.nacion.com/deportes/Costa-Rica-cantidad-liguistas-saprissistas_0_1320468079.html

Responda las siguientes preguntas:

1. ¿Qué quiere decir el periodista con el término *empate técnico* al inicio del reportaje?
2. ¿Por qué se menciona en el reportaje que existe igual cantidad de *liguistas* que de *saprissistas*, si los porcentajes de simpatizantes de cada uno de estos equipos presentan una diferencia de 2,6 puntos porcentuales?
3. ¿Qué significa el concepto de error de muestreo de la estimación que se utiliza al final del reportaje?

Análisis de la actividad 7

En el ejemplo anterior se plantean algunos conceptos que son citados regularmente por los medios de comunicación. El término *empate técnico* es muy empleado en encuestas políticas. En este caso particular significa que la diferencia observada en la muestra aleatoria respecto a los porcentajes de seguidores *liguistas* (35,9%) y *saprissistas* (33,3%) no resulta significativa para indicar que en la población total hay más *liguistas* que *saprissistas* (o viceversa). Esto ocurre porque dicha diferencia se encuentra dentro del margen del error de muestreo o error en la estimación de la encuesta aplicada. Como se explicará más adelante, el error de muestreo constituye la diferencia entre el valor de la población (desconocido) y la estimación que se genera por medio de la muestra.

Aunque no se incluye en el reportaje, la encuesta en mención tenía una probabilidad del 95% y el error máximo en la estimación era del 3,3%. Por esta razón, debido a que la diferencia observada entre los porcentajes de seguidores *liguistas* y *saprissistas* no supera este valor, entonces se menciona que existe un *empate técnico*.

La afirmación de que la afirmación de que “*En el país hay igual cantidad de liguistas que de saprissistas*” no necesariamente es cierta: lo que ocurre es que, mediante los resultados de la encuesta, no se pudo demostrar que existan diferencias entre los porcentajes de seguidores de ambos equipos en la población, lo que representa un concepto completamente distinto. Podría ocurrir que si se realizara un nuevo estudio con una muestra aleatoria más grande, se podrían encontrar diferencias en los porcentajes poblacionales a favor de uno u otro equipo.

Tampoco indica el estudio que la encuesta fue telefónica, por lo que la población a la que se hace referencia corresponde a la personas que tienen teléfono no privado y cuyo número está inscrito en el marco muestral empleado por los responsables del estudio.

La situación descrita anteriormente presenta un buen ejemplo introductorio para discutir el concepto de error aleatorio y otros tipos de errores presentes en los estudios estadísticos.

Errores en los análisis estadísticos.

En Estadística el concepto de error tiene una connotación muy particular, debido a que está incorporado a la mayoría de análisis inferenciales. En estos estudios se establecen estrategias para minimizar estos errores. Sin embargo, dentro de un análisis estadístico existen diferentes fuentes de error.

La principal clasificación que se puede realizar de los errores que pueden intervenir en un estudio estadístico consiste en diferenciar entre los **errores aleatorios** y los **errores no aleatorios** o **sesgos**.

Errores aleatorios

La diferencia generada entre el valor poblacional desconocido y la estimación que se realiza por medio de una muestra aleatoria se llama **error de muestreo**.

Por ejemplo, suponga que en una Dirección Regional de Costa Rica se desea conocer el porcentaje de estudiantes que abandona el colegio por problemas económicos en su hogar durante el primer año (suponga que se representa con P este porcentaje). El valor de P es desconocido, por ello se selecciona una muestra aleatoria de 200 estudiantes que han abandonado el colegio en los últimos años, se determina que 80 lo han hecho por problemas económicos en su hogar, lo que representa el 40% de los estudiantes de la muestra aleatoria. Este 40% corresponde a una estimación del porcentaje real de estudiantes “ P ”. La diferencia entre el valor de P y el valor estimado del 40% se llama error de muestreo en la estimación o error en la estimación:

$$\text{Error en la estimación} = P - 0,40$$

Como no se conoce el valor real de P , tampoco se conoce la magnitud del error en la estimación. Sin embargo, una de las virtudes que tiene el muestreo aleatorio consiste en la capacidad para determinar con anterioridad, y sujeto a una determinada probabilidad, la magnitud máxima del error de muestreo que se podría cometer al realizar esta estimación. Normalmente los errores de muestreo se determinan con una alta probabilidad (superior o igual al 90%), con el propósito de que el error máximo de muestreo sea lo más preciso posible.

Para el ejemplo de arriba, si existe una alta probabilidad (90%) que el error de muestreo no supere el 5%, entonces se puede asegurar con esta misma probabilidad que la proporción P de estudiantes que abandona el colegio por problemas económicos está contenido en el intervalo que va de 35% a 45%.

El concepto de error de muestreo corresponde a los análisis inferenciales, su cálculo no está presente en los programas de estudio del Ministerio de Educación Pública debido a que se requieren elementos de probabilidad de un nivel superior al que se ha considerado conveniente incluir en dichos programas. No obstante, este término es muy utilizado por los medios de comunicación al hacer referencia a los resultados de encuestas públicas y de otros estudios estadísticos, por esta razón es esencial que los docentes puedan comprender su significado y aplicación a problemas concretos.

Sesgos

A diferencia de los errores aleatorios los sesgos no son aleatorios, sino sistemáticos y normalmente obedecen a malas prácticas de selección de los datos o de las estrategias de recolección. En este sentido existen diferentes tipos de sesgos:

Sesgos de selección

Se presentan cuando se ha cometido un error en la selección de las unidades estadísticas incorporadas en el estudio. En estos casos normalmente la muestra seleccionada no representa a la población en estudio, por lo que no es pertinente para el problema y termina generando conclusiones erróneas. Este fue el problema que se presentó en la actividad 5, cuando la profesora Raquel tomó una muestra de estudiantes de séptimo año y en realidad debió seleccionar estudiantes del ciclo diversificado. Muchas veces se presenta este sesgo cuando se pretende realizar inferencias hacia la población pero no se ha utilizado una muestra aleatoria.

Sesgos de medición

Estos sesgos están muy vinculados con la estrategia de recolección de información: observación, experimentación o interrogación (encuestas). Estos sesgos se pueden presentar de diferentes formas:

1. **Sesgo del observador o enumerado:** se refiere a los errores que pueden cometer la persona o personas encargadas de recabar los datos. La capacidad de observación varía entre una persona y otra. También entran en juego los prejuicios que estas personas puedan tener en relación con el tema (aunque no sea de forma intencionada), muchas veces se promueve inconscientemente que los resultados respalden sus creencias. Otra causa para este tipo de sesgos consiste en los descuidos de quienes recolectan los datos.
2. **Sesgo del instrumento:** existen diferentes tipos de instrumentos para recolectar información estadística, los cuales dependen del tipo de datos que se van a recolectar. Estos instrumentos van desde instrumentos simples de medición como termómetros, microscopios, cintas métricas, entre otros que se utilizan regularmente en experimentos; hasta los cuestionarios que se emplean en la aplicación de encuestas. Estos instrumentos deben ser **confiables**, lo que quiere decir que si se aplican dos o más veces en igualdad de condiciones los resultados deben ser similares.
3. **Otros sesgos:** además de los sesgos ya presentados, durante la etapa del manejo de recolección de datos y su organización o sistematización se pueden cometer otros errores que tienden a sesgar la información, por ejemplo: pérdida de información, agrupamiento indebido de datos, exceso de redondeo, entre muchos otros.

Análisis didáctico

En relación con el tema de muestreo, en noveno año (página 364 de los programas de estudio) se incluye la siguiente habilidad.

Noveno año	Habilidad
Muestras aleatorias	1. Identificar la importancia del azar en los procesos de muestreo estadístico.

El análisis planteado en esta sección supera en mucho lo que se pretende con esta habilidad. Sin embargo, es muy importante que los docentes tengan un mayor conocimiento de esta temática, que el que se asume en los programas de estudio, para que puedan diseñar actividades didácticas con mayor propiedad.

Dicha habilidad debe ser concebida con mucha amplitud, debido a que la identificación del azar en los procesos de muestreo estadístico implican discutir sobre la potestad que tienen las muestras aleatorias de inferir sobre la población, situación que no se presenta en las muestras no aleatorias, en la cuales el azar no interviene. Por esta razón, es fundamental que el docente proponga problemas que involucren tanto muestras aleatorias como no aleatorias. Considere el siguiente problema para ser implementado en subgrupos de al menos cinco estudiantes de noveno año:

Problema:

Entre los aficionados al fútbol existe una discusión sobre ¿cuál es la afición más grande en el territorio nacional?, la mayor disputa ha estado centrada entre los aficionados a la Liga Deportiva Alajuelense y al Deportivo Saprissa. Realice una consulta a 50 estudiantes del colegio (hombres y mujeres) para determinar cuál es el equipo de fútbol preferido. Una vez recabada la información, determine el porcentaje de aficionados a favor de cada uno de los equipos que hayan sido mencionados.

1. *¿Qué se concluye respecto al equipo con mayor favoritismo?*
2. *Compare los resultados con los que se generaron en los otros subgrupos. ¿Existen diferencias en los porcentajes? ¿A qué atribuye estas diferencias?*

Una vez concluida esta primera parte, resuelva lo siguiente:

Estudiantes de la Escuela de Estadística de la Universidad de Costa Rica realizaron un estudio en donde, entre otras cosas, estimaron el porcentaje de seguidores de los equipos de fútbol de la primera división del país. El periódico La Nación publicó la siguiente información:

En Costa Rica hay igual cantidad de liguistas y de saprissistas

POR CAROLINA RUIZ V. - Actualizado el 29 de enero de 2013 a: 12:00 a.m.

San José (Redacción). Empate técnico. En el país hay igual cantidad de liguistas que de saprissistas, reveló una encuesta de la Universidad de Costa Rica (UCR).

Luego de consultar a 842 personas, como parte de la Encuesta Actualidades 2012, dos estudiantes de la Escuela de Estadística de la UCR descubrieron que 35,9% son aficionados a la Liga Deportiva Alajuelense (LDA) y 33,3%, al Deportivo Saprissa.

Al tomar en cuenta el error de muestreo de la estimación, se estima el empate técnico entre ambas barras de seguidores.

La última encuesta de Unimer había señalado que Saprissa era el más popular en abril del año pasado; pero a inicios del 2012 esa misma empresa de encuestas había reportado un empate técnico donde la LDA era un poco más preferido que el club tibaseño.

Según el estudio de los estudiantes de la UCR, a ambos equipos les siguen Heredia (5,6%) y Cartago (2,3%), mientras que cerca del 20% no apoya a ninguno.



http://www.nacion.com/deportes/Costa-Rica-cantidad-liquistas-saprissistas_0_1320468079.html

Responda las siguientes preguntas:

1. ¿Qué quiere decir el periodista con el término empate técnico al inicio del reportaje?
2. ¿Por qué se menciona en el reportaje que existe igual cantidad de liguistas que de saprissistas, si los porcentajes de simpatizantes de cada uno de estos equipos presentan una diferencia de 2,6 puntos porcentuales?
3. ¿Qué significa el concepto de error de muestreo de la estimación que se utiliza al final del reportaje?
4. ¿Existen diferencias entre los resultados de la encuesta y los resultados obtenidos en la etapa previa? ¿A qué atribuye las diferencias?
5. ¿Cómo cree que seleccionaron la muestra de 842 personas, en el estudio?

Este problema tiene múltiples propósitos: en primer lugar, en cada subgrupo los estudiantes deben seleccionar 50 compañeros de la institución para consultarles sobre su equipo de fútbol preferido. Posiblemente ninguna de las muestras sea aleatoria sino que responda a la conveniencia de cada subgrupo. El docente debe estar atento para identificar las diferentes formas de selección de la muestra, en los subgrupos, para evidenciarlo en la discusión interactiva y comunicativa focalizando sobre diferencias y semejanzas de dicha selección.

Una vez recabada la información, deben determinar los porcentajes correspondientes a cada uno de los equipos y comparar los resultados con los obtenidos por los otros subgrupos. Se espera que existan diferencias entre los resultados de los diferentes subgrupos.

Al plantear la segunda actividad, los estudiantes se enfrentan a conceptos técnicos desde el punto de vista estadístico que posiblemente han escuchado mencionar en algún momento, pero que posiblemente no han comprendido o no les han parecido relevantes. Pese a ello, ahora deben interpretar el reportaje periodístico y relacionar lo que está allí impreso con los resultados de la actividad previa que realizaron. Se esperaría que se generen dudas relacionadas con la posibilidad de que la terminología empleado en el artículo también pueda aplicar al estudio que realizaron en colegio.

Este es un problema en el cual el contexto juega papel sustancial, pues dependiendo del lugar del país donde se encuentre la institución, las preferencias por los equipos de fútbol pueden cambiar y ser muy distintas a las que se señalan en el reportaje.

Producto de lo anterior, la etapa de clausura o cierre del problema plantea considerables retos para el docente. En la discusión con los estudiantes, deben involucrarse en cuenta todos los aspectos discutidos en esta sección sobre el muestreo aleatorio y no aleatorio, así como las estrategias básicas del muestreo aleatorio. Es significativo que los estudiantes puedan diferenciar entre técnicas de muestreo, desde una perspectiva simple y concisa. Ante la posibilidad de que hayan empleado un muestreo no aleatorio en la etapa previa (posiblemente un muestreo por conveniencia), se recomienda discutir al menos los siguientes aspectos:

1. Evidenciar que las diferencias generadas en los porcentajes entre los resultados de los subgrupos obedecen a las diversas muestras seleccionadas. En esta discusión, analizar la forma en que seleccionaron a los estudiantes que consultaron y el tipo de pregunta que realizaron. Este tema es valioso, porque podría ocurrir que algún grupo no haya tomado en cuenta aquellas personas que no tenían preferencia por algún equipo o que no les gusta el fútbol, lo cual es un sesgo de selección. También se deben analizar otros posibles errores que se hayan presentado en esta etapa.

2. En la segunda parte, es valioso mencionar todo lo que sí fue discutido en el análisis de la actividad 7. En primer lugar el docente debe rescatar el hecho de que la muestra utilizada por los estudiantes de la Universidad de Costa Rica para realizar el estudio, es una muestra aleatoria, en la cual todas las personas de la población correspondiente (personas con teléfono no privado e incluidas en el estudio) tenían probabilidad de ser seleccionadas.
3. Un aspecto de gran relevancia en este proceso consiste en el significado de los porcentajes poblacionales, es decir el porcentaje de aficionados de la población total que sigue a un equipo determinado. Este valor es desconocido, solamente se puede conocer si se realiza una consulta a toda la población, lo cual es muy difícil por el elevado costo económico, el tiempo que conlleva hacerlo y otros elementos que influyen para que este tipo de estudios no se lleven a cabo (con excepción de los censos de población y vivienda que se realizan cada década). Por ello, para estimar estos valores se realizan encuestas por muestreo aleatorio, las cuales permiten estimar estos porcentajes con un determinado margen de error. De allí que en el artículo se hable de *error en la estimación*. Por eso la importancia que tiene el azar en la selección de la muestra.
4. Es necesario enfatizar en que tanto los porcentajes tanto poblacionales como las estimaciones muestrales, corresponden a representaciones numéricas que tienen un significado en el contexto del problema, pues se vinculan con las preferencias por los equipos de fútbol.
5. Otro concepto que debe quedar suficientemente claro y que está relacionado con los dos anteriores es el de *empate técnico*, que se utiliza para indicar que dos porcentajes se encuentran dentro del margen de error de 3,3 puntos porcentuales. Cuando se comparan dos estimaciones y sus diferencias están dentro del margen de error, se dice que existe un empate técnico; lo que significa que no hay evidencia suficiente para decir que los valores que se están estimando sean diferentes. Por ejemplo, se determinó que el 35,9% de los aficionados son liguistas, el margen de error significa que existe una alta probabilidad que el porcentaje de aficionados liguistas estaría comprendido en el intervalo $35,9 - 3,3 = 32,6$ y $35,9 + 3,3 = 39,2$. Asimismo, el porcentaje poblacional de aficionados saprissistas estaría entre 30,3 y 36,6; del mismo modo se aplicaría para los demás equipos.
6. El docente debería hacer hincapié también en los errores presentes el lenguaje empleado en el reportaje, especialmente en lo relacionado con la frase "*En el país hay igual cantidad de liguistas que de saprissistas*", tal como fue discutido en el análisis de la actividad 7. De igual manera debe referirse al concepto de población cuando se utiliza una encuesta telefónica.

Elaboración de bases de datos

La información que se requiere para la resolución de un problema debe estar organizada de manera sencilla para proceder con los análisis estadísticos correspondientes.

Actividad 8.

Se realizó un estudio para determinar el grado de contaminación del agua en la cuenca alta del río Virilla, específicamente en la zona norte del cantón de Coronado. Para ello se seleccionaron aleatoriamente 10 sitios estratégicos, y se tomaron dos muestras de agua en cada sitio el día 15 de cada uno de los 12 meses del año 2014. Cada muestra de agua fue analizada en un laboratorio químico y se determinó el valor de las siguientes variables:

1. **Sitio de muestreo:** se establecieron 10 sitios de muestreo, numerados de 1 a 10.
2. **Número de muestreo:** se realizan dos muestreos por sitio, se asigna el 1 al primer muestreo y 2 al segundo muestreo.
3. **Mes:** mes del año 2014 en que se realizó el muestreo.
4. **Porc.SO:** porcentaje de sólidos totales encontrados en el agua.
5. **DBO:** demanda bioquímica de oxígeno medida en mg/litro.
6. **Colif-fec:** coliformes fecales encontrados en el agua medidos en 200 microorg./100 ml.
7. **PH:** índice de acidez o alcalinidad del agua. Toma valores en el intervalo $[0, 14]$ y determina la acidez de una sustancia por medio de la relación con los iones de hidrógeno presentes. Valores de pH por debajo de 7 indican que una sustancia es ácida, mientras que si está por encima de 7 indica que es básica (el valor 7 se considera neutro).
8. **E-coli:** cantidad de bacterias Escherichia coli presentes en el agua, se mide en NPM/100ml (Número más probable de bacterias en 100 mililitros).
9. **Índice:** corresponde al índice holandés de la calidad del agua, es un valor numérico calculado a partir de la suma de valores vinculados con las variables más representativas de la calidad del agua, toma valores enteros entre 3 y 15. Entre más alto sea el índice menor es la calidad
10. **Calidad:** constituye la clasificación de la calidad del agua.

1: Sin contaminación.	2: Contaminación incipiente.
3: Contaminación moderada.	4: Contaminación severa.
5: Contaminación muy severa.	

Tomando en cuenta estos elementos, responda lo siguiente:

1. ¿Cuál es la unidad estadística del estudio?
2. ¿Cuántas unidades estadísticas incluyó el estudio?
3. ¿Cuántos datos se generaron en total?
4. *Elabore un esquema tabular por medio del cual usted organizaría estos datos con el fin de que puedan ser utilizados de una manera eficiente para llevar a cabo análisis estadísticos.*

Análisis de la actividad 8

La unidad estadística se concibe como la unidad básica de observación, esto quiere decir que es el sujeto u objeto que proporciona los datos para cada una de las variables consideradas en el estudio. Cuando se realiza una investigación que involucra personas, animales u objetos simples, es sencillo identificar las unidades estadísticas. Sin embargo, lo que se observa en el problema es el agua que fluye por el río, esto trae un componente de complejidad a las personas que realizan la investigación. Ellos toman muestras de agua en diferentes puntos de la cuenca alta del río Virilla, por ende cada muestra de agua debe verse como una unidad estadística.

En el planteamiento del problema se indica que se recolectaron dos muestras de agua en 10 sitios estratégicos, el día 15 de cada mes durante los 12 meses del año 2014. Esto da como resultado un total de $2 \cdot 10 \cdot 12 = 240$ registros, que equivalen al análisis de 240 unidades estadísticas. Para determinar la cantidad de datos, se debe considerar que cada muestra de agua (unidad estadística) fue analizada y se determinaron valores para 10 variables, entonces el estudio incluyó $10 \cdot 240 = 2400$ datos.

El problema general que se tiene al manejar tanta información consiste en identificar una forma simple que pueda resumir estos 2400 datos de manera que por una lado puedan ser accesibles para la persona que realiza el estudio, y al mismo tiempo puedan utilizarse en forma ágil para los análisis requeridos. La mejor forma de organizar este tipo de información, consiste en la elaboración de una base de datos, manteniendo el siguiente esquema:

<i>Número de registro</i>	<i>Sitio</i>	<i>Muestreo</i>	<i>Mes</i>	<i>Porc.SO</i>	<i>DBO</i>	<i>Coliformes fecales</i>	<i>PH</i>	<i>E-coli</i>	<i>Índice</i>	<i>Calidad</i>
1	1	1	1							
2	1	2	1							
3	2	1	1							
4	2	2	1							
5	3	1	1							
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
240	10	2	12							

Nota: no se incluyen los valores de las otras variables porque no se tienen.

Esta forma de organizar la información con variables que han sido codificadas previamente permite que la persona que realiza el estudio pueda tener acceso a los datos de una manera sencilla. Además, puede ser

utilizada por programas de computadora, tales como hojas de cálculo o incluso programas estadísticos, para sistematizar y aplicar técnicas estadísticas de forma accesible.

La organización de la información por medio de una base de datos similar a la anterior es una herramienta de fundamental importancia para los análisis que se vayan a realizar a partir de ellos. Toda persona que necesita realizar estudios estadísticos, sean simples o complejos, debe estar en capacidad de organizar la información en una base de datos que sea de fácil comprensión e identificación. En este proceso, tal como se indicó previamente, muchas de las variables incluidas en un estudio deben ser codificadas de modo que la organización sea lo más simple posible.

Las variables o características cualitativas normalmente requieren de una codificación que consiste en asignar valores numéricos a las categorías. En las variables cualitativas nominales esta codificación no tiene un orden establecido, es decir es indiferente el número que se asigne a cada categoría. Por el contrario, en las características cualitativas ordinales existe un orden lógico en las categorías, lo cual implica que los números asignados a ellas tienen también un orden (de mayor a menor o viceversa).

Las variables cuantitativas, sean discretas o continuas, normalmente no requieren de codificación debido a que se incluye en la base de datos el valor numérico correspondiente. Sin embargo, en algunas ocasiones, se procede a agrupar en clases las variables cuantitativas, lo que las convierte en variables cualitativas ordinales.

En el análisis de la actividad 8, la variable cualitativa “sitio” es nominal, pues no se especifica que los sitios deban clasificarse en un orden particular. Por su parte, las variables cualitativas “muestreo”, “mes” y “calidad” son ordinales, pues sus categorías tienen un orden claramente establecido. Las restantes variables son cuantitativas, todas ellas pueden ser catalogadas como continuas (se requiere de técnicas de medición para obtener su valor), con excepción de “índice” que corresponde a la suma de valores enteros, lo que la cataloga como variable discreta.

Análisis didáctico

Las habilidades vinculadas con las bases de datos no aparecen directamente dentro de la temática de secundaria, pero sí aparecen en I y II ciclo. No obstante, su uso es transversal en todos los niveles educativos, pues su implementación se relaciona directamente con el logro de muchas habilidades. Por esta razón, se puede decir que la elaboración y comprensión de las bases de datos constituyen una herramienta clave dentro de los estudios estadísticos que se promueven en toda la secundaria.

Seguidamente se plantea un problema estadístico que evidencia el rol que juegan las bases de datos en su resolución.

Problema:

En un estudio realizado por la Licenciada Alejandra Gamboa de la Escuela de Química de la Universidad Nacional, se pretendió determinar el grado de contaminación de la cuenca alta del río Tibás en las montañas de San Rafael de Heredia. El estudio incluyó el análisis de diferentes características de las muestras de agua tomadas en seis sitios seleccionados aleatoriamente en un período de 10 meses del año 2003. Algunas de las variables observadas fueron:

1. *Época: corresponde a la clasificación de los meses según época seca o lluviosa.*
 1: época seca 2: época lluviosa
2. *Sitio: representa el sitio del río donde se seleccionó la muestra de agua. Se seleccionaron seis sitios estratégicos donde se tomaron las muestras, la numeración de los sitios no tiene un orden preestablecido.*
3. *%SO: porcentaje de sólidos totales encontrados en el agua.*
4. *DBO: demanda bioquímica de oxígeno medida en mg/litro.*
5. *Colif-fec: coliformes fecales encontrados en el agua medidos en 200 microorg./100 ml*
6. *PH: índice de acidez o alcalinidad del agua. Toma valores en el intervalo [0,14] y determina la acidez de una sustancia, por medio de la relación con los iones de hidrógeno presentes. Valores de pH por debajo de 7 indican que una sustancia es ácida, mientras que si está por encima de 7 indica que es básica (el valor 7 se considera neutro).*
7. *Índice: medición de la calidad del agua según índice holandés, corresponde a la suma de puntos del comportamiento de algunas variables, normalmente se redondea a valores enteros entre 3 y 15. Entre más alto sea el índice menor es la calidad*
8. *Calidad: calidad del agua.*
 1: sin contaminación 2: contaminación incipiente 3: contaminación moderada

Seguidamente se presenta una muestra de 30 mediciones generadas en este estudio, incluyendo en la base de datos las variables época y sitio de muestreo, para contextualizar el análisis.

Época	Sitio	Sub Total %SO	DBO mg/L	Coliformes Fecales	PH	Índice de calidad	Calidad
1	3	16,65	1,0	5,94	9,18	4	2
1	2	16,66	1,0	6,66	8,97	4	2
1	2	16,66	1,2	5,77	8,73	3	1
1	5	16,66	1,5	8,77	9,45	3	1
1	3	16,63	1,8	5,87	9,05	4	2
1	4	16,61	1,9	7,52	8,80	4	2
1	6	16,66	2,5	11,40	8,77	4	2
1	1	16,66	3,0	7,47	8,73	4	2
1	1	16,66	3,7	13,78	8,28	5	2
1	1	16,65	4,2	9,82	9,15	4	2
1	4	16,66	5,5	12,07	8,84	5	2
1	5	16,66	8,0	11,40	9,09	6	2
1	4	16,55	14,0	8,77	8,99	7	3
1	3	16,57	14,0	6,47	9,20	7	3
1	6	16,56	14,0	8,77	9,13	7	3
2	4	16,54	1,0	8,77	9,23	4	2

2	5	15,92	1,0	9,82	9,31	4	2
2	5	16,66	1,2	8,87	9,55	4	2
2	1	16,64	1,2	7,47	9,55	4	2
2	6	16,65	1,2	5,87	9,31	3	1
2	2	16,64	1,2	5,87	9,45	3	1
2	4	16,64	2,3	7,95	8,95	3	1
2	5	16,62	3,0	7,47	9,05	4	2
2	1	16,66	3,0	11,40	9,31	3	1
2	1	16,54	3,3	10,33	9,18	4	2
2	6	14,25	3,8	14,11	8,50	5	2
2	3	16,66	4,1	8,57	9,31	4	2
2	2	16,24	4,4	5,87	8,70	5	2
2	5	14,42	4,9	9,82	9,41	5	2
2	1	16,66	8,4	9,82	9,31	6	2

De acuerdo con los datos incluidos en la base anterior y sabiendo que la presencia de coliformes fecales en el agua pone en riesgo la salud de las personas. Algunos estudios indican que si la cantidad de coliformes fecales es menor de 7 unidades (200 microorg./100 ml), el riesgo para la salud es bajo, mientras que si esta cantidad está entre 7 y 100 unidades, se considera que existe un riesgo intermedio, valores superiores a las 100 unidades representan un alto riesgo para la salud. Utilice la información de la base de datos para determinar:

1. El porcentaje de muestras de agua para las cuales el riesgo para la salud es intermedio.
2. Los sitios para los cuales las muestras de agua recolectadas estaban menos contaminadas y más contaminados con coliformes fecales.
3. Determine si existe algún tipo de relación entre la demanda bioquímica de oxígeno y el índice de calidad del agua.

Tenga presente que la demanda bioquímica de oxígeno consiste en la cantidad de oxígeno requerido por las bacterias durante cinco días a 20 grados centígrados. Esta variable es utilizada para determinar la contaminación de aguas domésticas en términos de la cantidad de oxígeno descargado en cuerpos naturales en donde prevalecen condiciones aeróbicas. Esta prueba es una de las más pertinentes en el control de actividades contaminantes en ríos.

Este problema puede ser adaptado prácticamente a todos los niveles de la secundaria modificando las interrogantes. En los términos en que fue redactado aquí se ajusta al nivel de octavo año:

Octavo año	Habilidades específicas
<p>Frecuencia</p> <p>a.) Absoluta</p> <p>b.) Porcentual</p> <p>Representación</p> <p>c.) Tabular: cuadros de frecuencia absoluta y porcentual</p> <p>d.) Gráfica: barras, circulares, lineales y diagramas de puntos</p> <p>Medidas de posición</p> <p>e.) Media aritmética</p>	<p>Utilizar representaciones tabulares o gráficas con frecuencias absolutas o porcentuales, simples o comparativas.</p> <p>Caracterizar un grupo de datos utilizando medidas estadísticas de resumen: moda, media aritmética, máximo, mínimo y recorrido.</p>

A continuación se presenta la solución de manera detallada:

Para responder lo propuesto en el inciso 1 se debe determinar la cantidad de muestras de agua que presentaron una cantidad de coliformes fecales de siete o más unidades (200 microorg./100 ml). Al realizar el conteo se determina que 22 de las 30 muestras de agua, lo que equivale al 73,3% de ellas, presenta un riesgo intermedio para la salud de las personas.

Para dar respuesta a lo propuesto en el inciso 2 se requiere observar los niveles de contaminación con coliformes fecales y los sitios donde se recolectaron las muestras, y se puede determinar el promedio de coliformes en cada caso. Cabe resaltar que la utilización del promedio cobra sentido, porque el número de muestras no es el mismo para cada sitio.

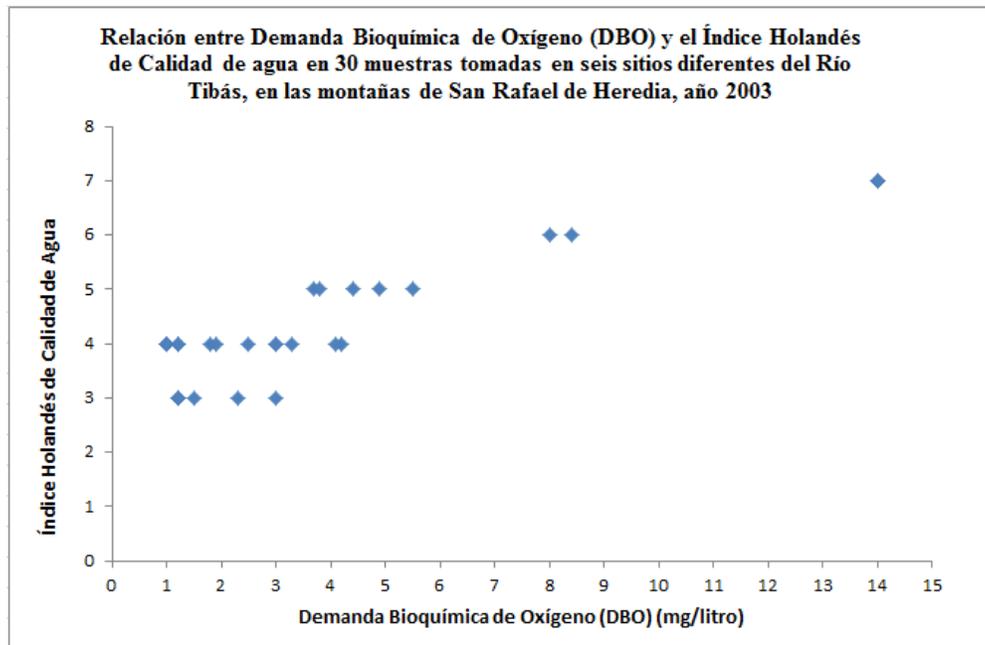
Sitio	Coliformes Fecales	Promedio	Sitio	Coliformes Fecales	Promedio
1	7,47	10,01	4	7,52	9,02
1	7,47		4	7,95	
1	9,82		4	8,77	
1	9,82		4	8,77	
1	10,33		4	12,07	
1	11,4		5	7,47	
1	13,78	5	8,77	9,36	
2	5,77	5	8,87		
2	5,87	5	9,82		
2	5,87	5	9,82		
2	6,66	5	11,4	10,04	
3	5,87	6	5,87		
3	5,94	6	8,77		
3	6,47	6	11,4		
3	8,57		6	14,11	

De acuerdo con lo anterior, se observa que las muestras de agua del sitio 2 son las menos contaminadas, todas las observaciones presentaron valores inferiores a siete. El segundo punto con muestras de agua menos contaminadas es el sitio 3; de las cuatros observaciones solamente en una de ellas se superó las siete unidades. Los promedios de coliformes por muestra de agua son bajos.

Por otro lado, las restantes muestras presentaron niveles de contaminación promedio superiores a siete y similares en los cuatro sitios. En los sitios 1, 4 y 5 todas las muestras observadas mostraron una cantidad de coliformes superior a siete. Por su parte el sitio 6 presentó un valor bajo, pero también un valor muy alto mostrando así mayor inestabilidad.

Se debe aclarar que este estudio se limita a las muestras observadas, no se están realizando generalizaciones a los sitios de donde se tomaron las muestras. Ni siquiera se está midiendo la variabilidad en cada sitio, simplemente se observa dónde fueron más estables las mediciones y dónde fueron más inestables (el concepto de desviación estándar y variancia se analiza en undécimo año).

Para el análisis del inciso 3 considere el siguiente gráfico:

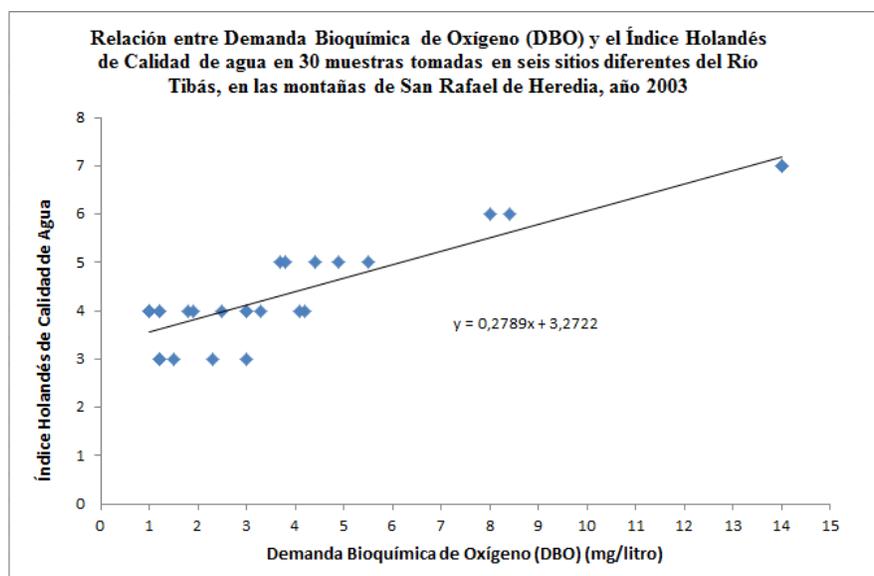


Fuente: Escuela de Química de la Universidad Nacional.

Por último, para determinar intuitivamente si existe relación entre la demanda bioquímica de oxígeno y el índice de calidad del agua, debido a que son dos variables cuantitativas, se puede recurrir a una representación gráfica. En estos casos, los diagramas de puntos constituyen una muy buena herramienta visual.

En el gráfico se observa que a medida que la demanda bioquímica de oxígeno (DBO) aumenta el índice de calidad de agua también aumenta (hay que recordar que valores del índice altos significan que el agua está más contaminada), lo que refleja que el agua es de menor calidad. La relación no es perfecta, debido a que otras variables intervienen también en la calidad del agua, por ejemplo la presencia de coliformes y otros.

Esta relación podría incluso ser aproximada por una función matemática, tal como se muestra a continuación:



Fuente: Escuela de Química de la Universidad Nacional.

Al igual que la relación se aproximó a una función lineal, también pudieron aproximarse con otras funciones más complejas. Aunque la encontrada no representa realmente la relación verdadera sino que es simplemente una aproximación, la pendiente de la recta es una aproximación a la razón de cambio en el índice de calidad del agua, por cada unidad que aumenta el DBO.

El problema planteado y su solución, permiten visualizar la importancia de las bases de datos en los análisis estadísticos, y al mismo tiempo incorporan elementos claves del proceso de enseñanza que son discutidos en los programas de estudio. Por ejemplo, puede notarse cómo su solución activa diferentes procesos matemáticos; los procesos representar y conectar son claramente identificables, debido a que cada una de las interrogantes al problema se analiza con una diferente forma de representación estadística de los datos. Además, se enlaza con Relaciones y Álgebra, desde la perspectiva de razones, proporciones y porcentajes, pero también en los conocimientos sobre relaciones entre variables. Otros procesos como razonar y argumentar y comunicar forman parte intrínseca del análisis y solución del problema.

Del mismo modo se puede evidenciar la presencia de los ejes disciplinares, debido a que el problema tiene un contexto muy particular, la información empleada es real y corresponde a un estudio realizado en la Universidad Nacional. Igualmente está referenciado el valor de la tecnología, no solamente para el manejo de la base de datos sino también para la elaboración de las representaciones gráficas.

Debido a que con el problema no se ha pretendido generar conocimiento nuevo, no es necesario entrar en definiciones (a no ser que algún estudiante consulte), sino más bien analizar diferentes elementos que fueron incorporados en la resolución del problema.

Entre los aspectos más relevantes que deberían ser analizados en esta etapa se tienen:

1. Uso de las bases de datos: se debe rescatar la importancia de la elaboración de bases de datos para sistematizar la información que se recolecta por diferentes técnicas estadísticas. El problema presentó un experimento químico sobre contaminación de aguas; si la información recabada en varios meses de estudio no hubiera sido resumida de esa manera, resultaría sumamente difícil para el investigador proceder con el estudio.

2. Las bases de datos tienen la ventaja de que usted puede visualizar la información que le suministra cada unidad estadística para cada variable considerada en el estudio. De este modo es factible proceder con la revisión y corrección de algún eventual error. Permite reconocer también si las variables son cuantitativas discretas o continuas, o cualitativas nominales u ordinales.
3. Resulta fundamental que cada base de datos vaya acompañada de las definiciones de las variables, las unidades de medida que se utilizan para el caso cuantitativo y las categorías vinculadas con las variables cualitativas.
4. En esta etapa, el docente debe hacer hincapié en la importancia que conlleva una adecuada lectura de los datos, no solamente en cuanto a su magnitud sino también en su contexto. Antes de proceder a realizar cualquier análisis estadístico, quien realiza el estudio debe familiarizarse con la información e incluso buscar posibles errores en ella. Se pueden considerar algunas interrogantes que podrían responderse con el simple hecho de observar los datos. Por ejemplo, las primeras dos preguntas del problema requieren de poco análisis y más bien demandan una adecuada lectura de los datos.

También se requiere puntualizar en el uso de representaciones para visualizar el comportamiento de los datos. El uso de porcentajes, promedios, cuadros y gráficas ayudan a observar patrones en el comportamiento de los datos en relación con el problema que les dio origen.

Bibliografía

- Araujo, C. (s.f.) *La incultura Estadística en nuestra sociedad: Necesidad de revisar la enseñanza de la Estadística Básica*. Departamento de Estadísticas, Facultad de Matemáticas, Pontificia Universidad Católica de Chile. Santiago de Chile.
<http://www.mat.uc.cl/archivos/File/SOBRE.DOCENCIA/A01%20La%20Incultura%20Estadistica%20en%20Nuestra%20Sociedad.pdf>
- Arteaga P., Batanero C., Cañadas, G. y Contreras J. M. (2011). Las Tablas y Gráficos Estadísticos como Objetos Culturales. *Números: Revista de Didáctica de las Matemáticas*, (7)6, 55–67.
- Batanero, C. (2002). Los retos de la cultura estadística. *Yupana*, 1(04), 27- 37.
- Fischbein, E. (1987). *Intuition in science and mathematics*. Dordrecht: Reidel.
- Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- Ministerio de Educación Pública de Costa Rica (2012). *Programas de estudio de Matemáticas para la Educación General Básica y el Ciclo Diversificado*. San José, Costa Rica: autor.
- Ministerio de Educación Pública de Costa Rica, Proyecto Reforma de la Educación Matemática en Costa Rica (2012). *Curso bimodal para el Ciclo Diversificado: Enfoque de Resolución de problemas. Unidad didáctica Estadística*. San José, Costa Rica: autor.
- Ministerio de Educación Pública de Costa Rica, Proyecto Reforma de la Educación Matemática en Costa Rica (2012). *Curso bimodal para el Ciclo Diversificado: Enfoque de Resolución de problemas. Unidad didáctica Probabilidades*. San José, Costa Rica: autor.
- Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-265.

Créditos.

La enseñanza de la Estadística: mucho más que procedimientos y técnicas. Unidad didáctica. Primera parte es un recurso del *Curso bimodal de capacitación para docentes de la Educación Secundaria*. Una actividad del Proyecto Reforma de la Educación Matemática en Costa Rica.

Este proyecto del Ministerio de Educación Pública es apoyado por la Fundación Costa Rica - Estados Unidos de América para la Cooperación.

Autor de la unidad didáctica del curso

Edwin Chaves Esquivel

Revisores

Javier Barquero Rodríguez, Luis Hernández Solís, Erasmo López López, Keibel Ramírez Campos, Ángel Ruiz, Marianela Zumbado Castro.

Revisión Filológica

Julián Ruiz Blais

Edición gráfica

Keibel Ramírez Campos

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

Ángel Ruiz

La imagen de la portada es cortesía de <http://www.freedigitalphotos.net/>

Para referenciar este documento:

Ministerio de Educación Pública, Proyecto Reforma de la Educación Matemática en Costa Rica (2015). *La enseñanza de la Estadística: mucho más que procedimientos y técnicas. Unidad didáctica. Primera parte*. San José, Costa Rica: autor.



La enseñanza de la Estadística: mucho más que procedimientos y técnicas. Unidad didáctica. Primera parte, por Ministerio de Educación Pública de Costa Rica, Proyecto Reforma de la Educación Matemática en Costa Rica, se encuentra bajo una Licencia Creative Commons Atribución-CompartirIgual 3.0 Unported.