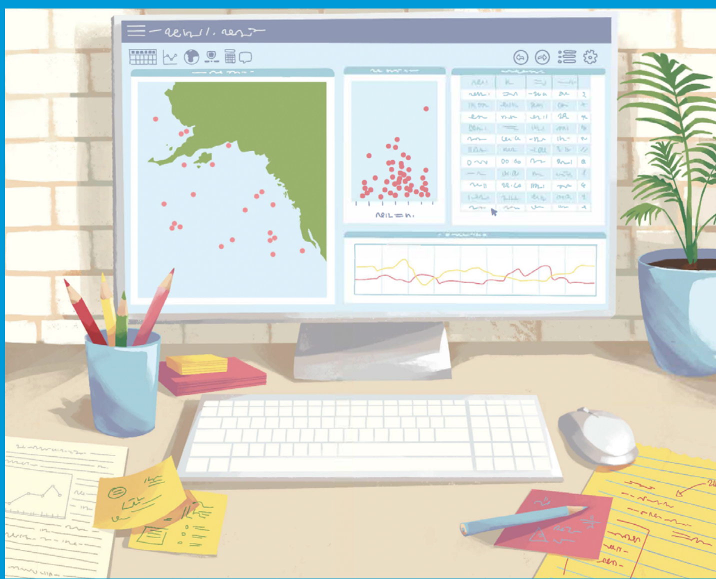
Pautas para la evaluación y la enseñanza en educación estadística de los 3 a los 18 años

(Informe GAISE II)

Un marco de trabajo para la educación en estadística y ciencia de datos

Anna Bargagliotti (copresidenta)
Christine Franklin (copresidenta)
Pip Arnold
Rob Gould
Sheri Johnson
Leticia Perez
Denise A. Spangler



NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

Con la colaboración de Fundación "la Caixa"

Pautas para la evaluación y la enseñanza en educación estadística de los 3 a los 18 años

(Informe GAISE II)

Un marco de trabajo para la educación en estadística y ciencia de datos

Título original:
Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)
A Framework for Statistics and Data Science Education

Pautas para la evaluación y la enseñanza en educación estadística de los 3 a los 18 años (Informe GAISE II)
Un marco de trabajo para la educación en estadística y ciencia de datos

Coordinación de la edición
Marta García Matos

Traducción
Luis J. Rodríguez-Muñiz (coordinador)
María Abella-Vázquez
Marlén Alonso-Castaño
Marián González-Rúa
Laura Muñiz-Rodríguez
Universidad de Oviedo

Revisión lingüística y maquetación
Montserrat Alberte y Toni Rovira (Lexware)

ÍNDICE

Agradecimientos

Un especial agradecimiento a Christine Franklin, Gary Kader, Denise Mewborn (Spangler), Jerry Moreno, Roxy Peck, Mikey Perry y Richard Scheaffer por su liderazgo y visión en el primer documento GAISE I.

El equipo de autores y autoras agradece sinceramente al ASA/NCTM Joint Board[1] la financiación de la redacción y la labor de producción del GAISE II; y, en especial, a Donna LaLonde y Rebecca Nichols, de la ASA, y a Dave Barnes y Jeff Shih, del NCTM, por su apoyo a lo largo de todo el proceso.

Gracias a Brenna Bastian por sus preciosas ilustraciones para la cubierta del informe y el ejemplo de las mariquitas del nivel A. También agradecemos el diseño y el trabajo de maquetación de Shirley E. M. Raybuck y Valerie Nirala.

Por último, el equipo de autores desea expresar su sentido reconocimiento a las veintidós personas que revisaron detenidamente este documento del GAISE II y nos proporcionaron una valiosa retroalimentación:

Gail Burrill, Rosemarie Callingham, Catherine Case, Michelle Dalrymple, Neville Davies, Ed Dickey, David Fluharty, Gary Kader, Donna LaLonde, Jerry Moreno, Rebecca Nichols, Regina Nuzzo, Roxy Peck, Jamis Perrett, Maxine Pfannkuch, Katherine Respress, Richard Schaeffer, Neil Sheldon, Josh Tabor, Dan Teague, Doug Tyson y Jane Watson.

El equipo traductor hace constar su agradecimiento a la Fundación "la Caixa" por financiar esta traducción del GAISE II, con un reconocimiento especial a Marta García Matos.

[1] N. de T.: Para facilitar la identificación de los organismos y asociaciones del contexto del documento original, hemos optado por no traducir su denominación y mantener la original en inglés.

Prefacio

En 2020, cuando se publica el informe *Pre-K–12 guidelines for assessment and instruction in Statistics education II: A framework for Statistics and Data Science education report (GAISE II)*,[2] la alfabetización estadística y en datos es más importante que nunca. Se insta a que las personas sinteticen información sobre muchas cuestiones que tienen un alcance global, como la pandemia del COVID-19, los cambios que sufre el planeta con condiciones climáticas extremas, los repuntes y las recesiones económicas, e importantes aspectos sociales, como el movimiento Black Lives Matter.[3] Los datos se presentan a través de visualizaciones (a veces interactivas y a veces no), informes de estudios científicos (como estudios médicos), artículos en prensa y páginas web.

Las demandas de alfabetización estadística nunca han sido mayores. Las personas que acaban la educación secundaria alfabetizadas estadísticamente deben ser capaces de evaluar las conclusiones y la legitimidad de los resultados informados, así como de formular sus propios análisis. Steve Levitt, coautor de los libros *Freakonomics*, abordó la necesidad de la alfabetización estadística y en datos en un *podcast* de Stephen Dubner (2019), en el que señalaba que tenemos la responsabilidad de preparar a nuestros hijos e hijas para el mundo que se van a encontrar: un mundo dirigido por los datos, y que una fluidez básica con los datos es un requisito no solo para la mayoría de los buenos empleos, sino también para hacerse camino en la vida en general, ya sea en términos de alfabetización financiera, de tomar buenas decisiones en cuanto a la propia salud o de saber a quién y qué creer.

Impulsados por la revolución digital, los datos son ahora fácilmente accesibles mediante métodos estadísticos y herramientas tecnológicas, lo que permite al alumnado obtener información y formular recomendaciones para gestionar problemas globales urgentes. Los datos pueden ser sumamente valiosos, pero solo si se utilizan con criterio y en un contexto adecuado.

Hoy en día, muchos sectores de la economía y la mayoría de los empleos dependen de habilidades con los datos. Se necesita contar con un buen sentido de los datos para leer las noticias con facilidad y participar en la sociedad como un miembro bien informado. Debido a esto, es esencial que todo el alumnado finalice la secundaria preparado para vivir y trabajar en un mundo dirigido por los datos. El informe GAISE II Pre-K–12 presenta una serie de recomendaciones para la alfabetización estadística a nivel escolar.

Resumen y objetivos

El documento *Guidelines for assessment and instruction in Statistics education (GAISE) report: A Pre-K–12 curriculum framework* (en adelante, GAISE I Pre-K–12), de Franklin et al., se publicó por primera vez en 2005, con pequeñas revisiones en 2007, junto con *Guidelines for assessment and instruction in Statistics education (GAISE). College report*. Este informe hacía una descripción general de algunas recomendaciones para el curso introductorio de estadística en educación superior y fue actualizado en 2016 (GAISE College Report ASA Revision Committee, 2016).

---

[2] N. de T.: Mantendremos la designación utilizada en algunos países, como los Estados Unidos, Canadá, Turquía, Filipinas, Australia y Ecuador. Esta nomenclatura hace referencia a todas las etapas educativas previas a la universitaria. Está formada por *Pre-K*, abreviación del término inglés *pre-kindergarten*, equivalente al 2.º ciclo de Educación Infantil (entre los 3 y los 5 años), y el número 12, que indica el último grado (entre los 17 y los 19 años); por tanto, Pre-K–12 cubre las etapas educativas entre los 3 y los 19 años. Su equivalente en España sería Educación Infantil (de 3 a 6 años), Educación Primaria (de 6 a 12 años) y Educación Secundaria (de 12 a 18 años). Asimismo, esta última etapa de secundaria está dividida en una etapa obligatoria (ESO) hasta los 16 años, y dos años más (Bachillerato) necesarios para poder acceder a estudios universitarios. En México equivaldría a Educación Inicial o Preescolar (de 3 a 5 años), Educación Primaria (de 6 a 12 años), Educación Secundaria (de 12 a 15 años) y Educación Media Superior o Preparatoria/Bachillerato (de 15 a 18 años). En Chile, la equivalencia sería Educación Parvularia (de 3 a 5 años), Educación Básica (de 6 a 13 años) y Educación Media (de 14 a 17 años). En Colombia, Pre-K–12 sería equivalente a Educación Preescolar (de 3 a 5 años), Educación Básica Primaria (de 6 a 10 años), Educación Básica Secundaria (de 11 a 14 años) y Educación Media (de 15 a 17 años). En Argentina, Nivel Inicial (Sala de 3, 4 y 5 años), Nivel Primario (de 6 a 12 años aprox.) y Nivel Secundario (de 13 a 17/18 años).

[3] N. de T.: *Black Lives Matter* es un movimiento internacional antirracista cuyo nombre podría traducirse literalmente como «las vidas de las personas negras importan».

El informe GAISE I Pre-K–12 se elaboró para ampliar los estándares estadísticos de *Principles and standards for school Mathematics* del National Council of Teachers of Mathematics (NCTM), publicado en el año 2000, y como continuación del documento *The mathematical education of teachers* (MET) de la Conference Board of Mathematical Sciences (CBMS). El GAISE I Pre-K–12 fue un informe seminal y visionario que abogó por la necesidad de la alfabetización estadística y en datos desde los primeros cursos escolares. Proporcionó un marco de recomendaciones para desarrollar las habilidades básicas de los estudiantes en razonamiento estadístico en tres niveles a lo largo de los años escolares, descritos como niveles A, B y C. Estos niveles se mantienen en el GAISE II y, en líneas generales, equivalen a primaria, secundaria y bachillerato. La progresión a través de los niveles secuenciales del GAISE Pre-K–12 (tanto I como II) es la esperada para quien aspire a alcanzar la alfabetización estadística, independientemente de su edad.

Desde su publicación inicial, el GAISE I ha tenido un impacto significativo en la inclusión de estándares estadísticos a nivel estatal y nacional en los Estados Unidos y en el resto del mundo. El informe ha sido utilizado internacionalmente como punto de referencia para la educación estadística a nivel escolar, y se dispone de una traducción al español (Franklin et al., 2018). En el momento de la redacción del GAISE II, el GAISE I contaba con más de 790 citas en trabajos académicos, según Google Scholar. Ha sido referenciado en numerosos proyectos financiados por la National Science Foundation y en los informes de otras organizaciones profesionales de educación en STEM (*science*, *technology*, *engineering*, *mathematics*). El GAISE I también ha influido en el desarrollo de estándares en los Estados Unidos y en la redacción del documento *Statistical education of teachers* (SET; Franklin et al., 2015), publicado por la ASA, en el que se hacen recomendaciones para la preparación de docentes de primaria y de secundaria en estadística.

El GAISE I se centraba fundamentalmente en datos tradicionales de variables cuantitativas y cualitativas y en estudios que utilizaban pequeños conjuntos de datos de muestras de una población. Quince años más tarde, los tipos de datos se han diversificado más allá de la clasificación en cuantitativos y cualitativos, lo que ha comportado la necesidad de adquirir competencias estadísticas diferentes y, con frecuencia, de última generación. Hoy, por ejemplo, los datos incluyen textos publicados en redes sociales o colecciones altamente estructuradas (o no estructuradas) de imágenes, sonidos o vídeos. Los datos son extremadamente numerosos y están fácilmente disponibles. Los datos son multidimensionales. Con frecuencia, los datos se representan y se visualizan también de manera multidimensional e interactiva, de modo que se presentan simultáneamente muchas variables.

El GAISE II incorpora las nuevas habilidades necesarias para dar sentido a los datos en la actualidad, sin perder el espíritu del GAISE I, y pone de relieve:

1. La importancia de plantear preguntas durante el proceso de resolución de problemas estadísticos (formular una pregunta de investigación estadística, recoger o considerar datos, analizarlos e interpretar los resultados), y cómo este proceso se mantiene a la vanguardia del razonamiento estadístico en todos los estudios que incluyen datos.
2. La consideración de diferentes tipos de datos y variables; la importancia de planificar cuidadosamente cómo recoger los datos o cómo sopesarlos para ayudar a responder preguntas de investigación estadística; y el proceso de recogida, limpieza, examen y análisis de los datos.
3. La inclusión del pensamiento multivariante a lo largo de todos los niveles educativos Pre-K–12.
4. El papel del pensamiento probabilístico en la cuantificación de la aleatoriedad a lo largo de todos los niveles.
5. El reconocimiento de que la práctica estadística moderna está entrelazada con la tecnología, y la importancia de incorporar la tecnología siempre que sea posible.
6. La relevancia de comunicar de manera clara y precisa la información estadística.
7. El papel de la evaluación a nivel escolar, especialmente de los ítems que miden la comprensión conceptual y que requieren un razonamiento estadístico en el que intervenga el proceso de resolución de problemas estadísticos.

Un futuro dirigido por los datos

Como se indica más adelante:

> Los datos se usan para contar una historia. Las personas que se dedican a la estadística ven el mundo a través de los datos —los datos proporcionan modelos de la realidad—. El pensamiento estadístico y el proceso de resolución de problemas estadísticos son fundamentales para explorar todos los datos.

El GAISE II presenta una visión en la que cada persona actúa con seguridad al razonar estadísticamente, dar sentido a los datos y saber cómo y cuándo aplicar un sano escepticismo a la información obtenida de los datos. Aquí se presenta un marco de trabajo de conceptos esenciales y 22 ejemplos a lo largo de tres niveles de desarrollo de habilidades. Este marco de trabajo sirve de apoyo a todo el alumnado mientras este aprende a apreciar el papel fundamental del razonamiento estadístico y la ciencia de datos, y adquiere la habilidad vital esencial de la alfabetización en datos.

Quienes escribimos este documento apreciamos la oportunidad, el apoyo y el respaldo del Consejo de Dirección de la American Statistical Association (ASA) y del National Council of Teachers of Mathematics (NCTM) para la mejora del documento GAISE I Pre-K–12. Nuestra esperanza es que el documento GAISE II Pre-K–12 enriquezca su trabajo y fomente el objetivo final: la alfabetización estadística para todo el mundo.

Anna Bargagliotti (copresidenta)
Christine Franklin (copresidenta)
Pip Arnold
Rob Gould
Sheri Johnson
Leticia Perez
Denise A. Spangler

Equipo traductor:
Luis J. Rodríguez-Muñiz (coordinador)
María Abella-Vázquez
Marlén Alonso-Castaño
Marián González-Rúa
Laura Muñiz-Rodríguez
Universidad de Oviedo

INTRODUCCIÓN

Nuestras vidas están enormemente influidas por los datos. Todo el mundo debería ser capaz de emplear un razonamiento estadístico sólido para tomar decisiones inteligentes basadas en la evidencia. La alfabetización estadística es necesaria para tener éxito en el trabajo, mantenernos informados sobre sucesos actuales y estar preparados para una vida sana, feliz y productiva.

Cada mañana, los periódicos y otros medios de comunicación nos muestran información estadística sobre temas que van desde la economía hasta la educación, desde el cine hasta el deporte, desde la comida hasta la medicina y desde la opinión pública hasta el comportamiento social. Esa información guía las decisiones en nuestra vida personal y nos capacita para afrontar nuestras responsabilidades como miembros de una comunidad y una sociedad. La alfabetización estadística es un requisito para abrirnos camino en el mundo de hoy.

Los datos también pueden ser útiles en nuestras vidas personales. Llevar puesta una pulsera inteligente de actividad nos permite monitorizar los pasos, la frecuencia cardiaca y otras estadísticas relacionadas con la condición física; y ello puede motivar estilos de vida más saludables. Si consideramos mudarnos a otra zona, podríamos tomar decisiones basadas en estadísticas sobre el coste de la vida o el clima local.

Es posible que las personas alfabetizadas estadísticamente tengan la oportunidad de progresar en sus carreras profesionales y obtener puestos de trabajo más gratificantes y de mayor desafío. Los directivos de empresas podrán recibir información cuantitativa sobre presupuestos, suministros, detalles de producción, demandas de mercado, previsión de ventas o carga de trabajo. El profesorado podrá hacer frente a estadísticas de educación relativas al desempeño del alumnado o al suyo propio. En el campo de la medicina, los científicos deben entender los resultados estadísticos de los experimentos llevados a cabo para probar la efectividad y la seguridad de los tratamientos médicos. Los profesionales de las fuerzas del orden dependen de las estadísticas de criminalidad.

La alfabetización estadística implica contar con una buena dosis de escepticismo sobre las conclusiones basadas en datos. Cualquier persona graduada en educación secundaria que esté alfabetizada estadísticamente será capaz de evaluar conclusiones a partir de datos y juzgar la legitimidad de los resultados obtenidos. El informe *The new foundational skills of the digital economy* del Business Higher Education Forum (Markow et al., 2018, p. 18) reconoce y afirma que, en algún momento del futuro, muchos de los altos niveles de competencia que actualmente parecen confinados a las cotas altas de la economía digital, o a organizaciones más grandes y complejas, se convertirán en la norma entre quienes buscan trabajo y quienes ya trabajan, así como en los propios lugares de trabajo.

El objetivo final: alfabetización estadística para todo el mundo

Desarrollar competencias estadísticas sólidas exige tiempo, y perfeccionarlas hasta el nivel necesario en el mundo moderno no es posible en un único curso de secundaria. La mejor manera de ayudar a las personas a lograr una alfabetización estadística es comenzando su educación en esta disciplina desde los niveles de educación infantil para continuar reforzando e incrementando sus competencias estadísticas a lo largo de la educación primaria y secundaria.

Este documento expone un marco curricular para los programas educativos de las etapas Pre-K–12, diseñado para ayudar al alumnado a alcanzar la alfabetización en datos y a estar alfabetizado estadísticamente. El marco de trabajo y las posteriores secciones del presente informe recomiendan estrategias curriculares y de implementación que abarcan la educación estadística en todos los niveles Pre-K–12.

El documento proporciona un marco curricular para las tres etapas educativas, pero el itinerario hacia la alfabetización estadística presentado aquí es apropiado para cualquier rango de edad. De hecho, está pensado para cualquier persona que trate de mejorar su alfabetización estadística. A lo largo de este informe nos referiremos a estas personas como «el alumnado».

El porqué de las *Pautas para la evaluación y la enseñanza en educación estadística (GAISE) II*. Visión general y objetivos del GAISE II

El GAISE I se publicó por primera vez en 2005 y se revisó, con pocos cambios, en 2007 (véase la figura 1). Fue un documento innovador. En él se defendía la alfabetización estadística a nivel escolar y se explicaba cómo el pensamiento matemático y el estadístico, aunque relacionados, son distintos. Las competencias matemáticas son necesarias, y el pensamiento estadístico y el matemático deben trabajar conjuntamente al analizar los datos.

Desde que se publicó el GAISE I, las concepciones tradicionales acerca de los datos han cambiado. Los datos ya no son simplemente números en un contexto, clasificados de forma cuantitativa o cualitativa,[4] generalmente almacenados en hojas de cálculo estáticas. Hoy, los datos pueden ser también colecciones dinámicas, complejas y muy estructuradas (o desestructuradas) de imágenes o sonidos. Los conjuntos de datos son amplios y están al alcance de todos.

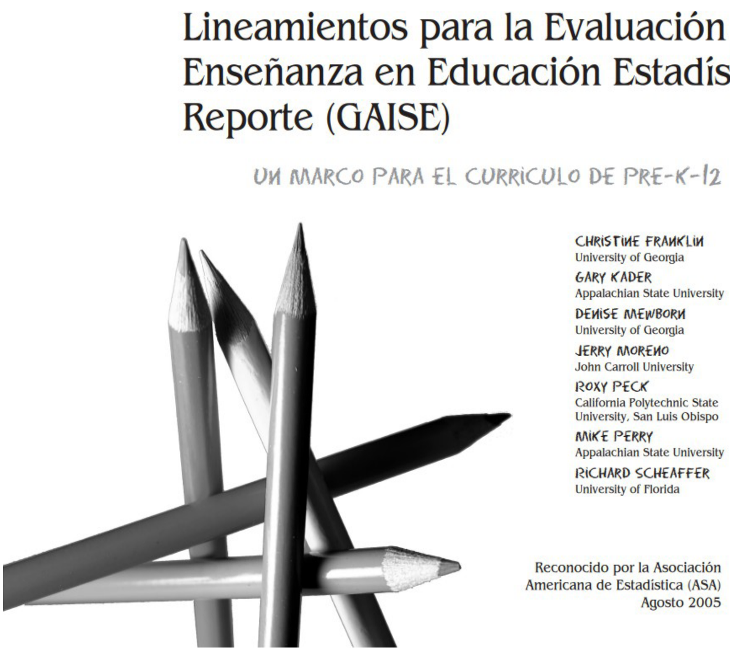


Figura 1. El GAISE I se publicó en 2005.

El GAISE II incluye ejemplos que tienen que ver con datos no tradicionales y multivariantes presentados a lo largo de todo el currículo de Pre-K–12. El alumnado tiene que desarrollar la alfabetización estadística para dar sentido a la inmensa cantidad de datos que lo rodean en su día a día. Muchos de estos datos son generados por el propio alumnado a través de sus redes sociales, cuando utilizan sistemas de geolocalización (GPS), etc. Deben ser conscientes de cómo se almacenan estos datos y cómo los utilizan las organizaciones que los recopilan. Asimismo, tienen que entender por qué nuestra sociedad necesita medidas y políticas de seguridad para prevenir la mala praxis y el uso poco ético de estos datos.

El alumnado debe empezar a manejar datos desde edades tempranas, tanto si trabaja con pequeños conjuntos de datos como con conjuntos grandes y desorganizados, con datos tradicionales o no, como textos o imágenes. La mayoría de los trabajos del futuro requerirán conocimientos de estadística y de análisis de datos.

[4] N. de T.: El original inglés es «classified as quantitative or categorical». En español, para referirse a «categorical», se puede hablar de *variables categóricas* o *variables cualitativas*, siendo aceptables ambas formas. Sin embargo, resulta más frecuente el término *variable cualitativa*, que es el utilizado en esta traducción. Habitualmente, los valores de una variable cualitativa o categórica se denominan *categorías*, término que utilizamos en el documento.

Ser capaz de razonar estadísticamente es esencial en todas las ramas de conocimiento y profesiones. Muchas disciplinas, como las ciencias, incluyen ahora la estadística en sus estándares, por ejemplo, los Estándares Científicos para las Próximas Generaciones (NGSS, por su sigla en inglés; National Research Council, 2013). Varios de los nuevos ejemplos del GAISE II usan conjuntos de datos científicos y tratan temas abordados en estos estándares, para ilustrar cómo el razonamiento estadístico es un componente integral de las investigaciones científicas.

Aunque en los últimos 50 años se ha avanzado en lo relativo a la formación del alumnado, aún queda mucho trabajo necesario para el futuro, dado que la estadística es una disciplina en continuo desarrollo. Entre los documentos importantes posteriores a la publicación del GAISE I que promueven la alfabetización estadística para todos los niveles educativos Pre-K–12 figuran *Statistical education of teachers* (Franklin et al., 2015) y los manuales *Catalyzing change*, con el subtítulo *Initiating critical conversations*, para primaria, secundaria y bachillerato (Bush et al., 2000; Graham et al., 2018; Huinker et al., 2020), publicados por el NCTM. Para más información sobre la historia de la estadística en Pre-K–12, se puede consultar el capítulo 9 del SET.

El GAISE II conserva el espíritu del informe GAISE I. El proceso de resolución de problemas estadísticos definido en el GAISE I sigue siendo la base y el núcleo del razonamiento estadístico y de todo lo que tiene que ver con dar sentido a los datos. Se define como un proceso compuesto de cuatro pasos: 1) formulación de una pregunta de investigación estadística; 2) recogida o consideración de los datos; 3) análisis de los datos; y 4) interpretación de los resultados. Los tres niveles A, B y C (que, en líneas generales, coinciden con primaria, secundaria y bachillerato) se mantienen de forma consistente en los informes GAISE I y GAISE II.

Impulsado por la sobreabundancia de datos disponibles en el mundo actual, el proceso de resolución de problemas estadísticos no solo mantiene su importancia, sino que adquiere mayor relevancia a la hora de extraer conclusiones a partir de los datos. Esto incluye reconocer las representaciones gráficas engañosas y las limitaciones de los conjuntos de datos, por muy grandes que sean, que se usan para responder las preguntas estadísticas de investigación. Para ver algunos ejemplos de estadísticas engañosas y de limitaciones de los conjuntos de datos, se puede consultar el artículo de Dominus (2017), publicado en *The New York Times Magazine*, o el de Kuiper (2010) sobre la incorporación de la experiencia investigadora a un curso inicial de estadística universitaria.

Los estudios estadísticamente sólidos deben ser reproducibles. Cualquier estudio que se repita con muestras similares debe llevar a conclusiones similares, y los diferentes métodos estadísticos que se apliquen de nuevo a los mismos datos deben dar resultados consistentes.

Tal y como se enfatiza en el primer informe GAISE: «La estadística es una disciplina metodológica. No existe por sí misma, sino que ofrece a otros campos de estudio un conjunto coherente de ideas y herramientas para el tratamiento de datos. La necesidad para tal disciplina surge de la *omnipresencia de la variabilidad*» (Cobb y Moore, 1997, p. 801, como se cita en Franklin et al., 2018). El pensamiento estadístico, en gran medida, debe lidiar con la omnipresencia de la variabilidad en los datos (p. ej., variabilidad dentro de un grupo, variabilidad entre grupos, variabilidad de un estadístico[5] de unas muestras a otras). La resolución de problemas estadísticos y la toma de decisiones dependen de la comprensión, la explicación y la cuantificación de la variabilidad de los datos dentro de un contexto dado. «La estadística requiere una forma diferente de pensar, porque *los datos no son solo números, sino números en un contexto*. En matemáticas, el contexto oscurece la estructura. En análisis de datos, el contexto proporciona significado» (Cobb y Moore, 1997, p. 801, como se cita en Franklin et al., 2018).

[5] N. de T.: El término inglés *statistic* puede ser traducido como *estadístico* (masculino) o *estadística* (femenino). Los significados de ambos términos son sinónimos a menudo, pero, en ocasiones, presentan algunos matices. En la traducción usaremos *estadístico* para referirnos a la función de valores obtenidos de una muestra aleatoria que se utiliza para inferir propiedades de una población; es decir, es una función cuyos valores son aleatorios y que, para cada muestra, toma unos valores concretos. Por su parte, *estadística* se suele utilizar para denotar un conjunto de medidas estadísticas que se corresponden con esos valores concretos del estadístico; es decir, no es una función, sino un valor o un conjunto de valores numéricos.

Para resaltar la importancia de la alfabetización estadística hoy en día, consideremos el gráfico de la figura 2, creado en 2014 y actualizado en 2017 por Douglas A. Webber, y publicado por el *New York Times* (The Learning Network, 2018).

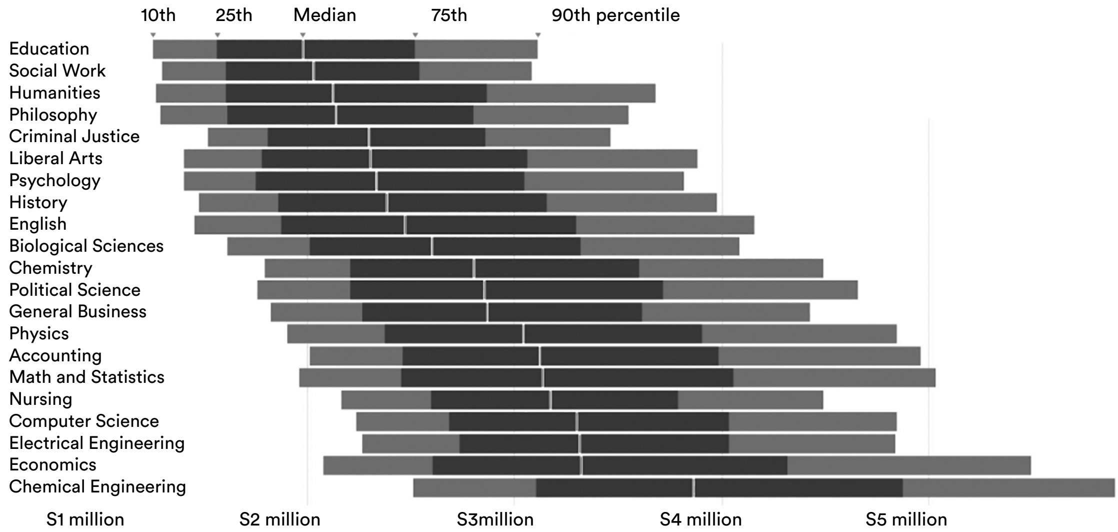


Figura 2. Distribuciones de los ingresos previstos para distintas profesiones

El gráfico de la figura 2 muestra las distribuciones de los ingresos previstos para varias profesiones. Las distribuciones se representan con barras divididas en segmentos por los percentiles 10, 25, 50, 75 y 90 (medidas de posición para cada distribución). Tanto el contexto como la variabilidad de las diferentes distribuciones son importantes a la hora de interpretar este gráfico. Por ejemplo, podemos apreciar que la educación es una de las profesiones peor remuneradas si la comparamos con la ingeniería química, pero hay también mayor variabilidad en los salarios de los ingenieros químicos (véase la longitud de los segmentos dentro de las barras). El contexto es igualmente relevante. Puede apreciarse que esta representación gráfica proporciona escasa información acerca de cómo se obtuvieron los datos. Es importante preguntarse cómo se obtuvieron: ¿son datos primarios recogidos por los investigadores o son datos secundarios puestos a disposición de los investigadores?

Son estas características (el enfoque de la variabilidad de los datos, la importancia del contexto asociado a los datos y el cuestionamiento de los datos) las que diferencian a la estadística de otras ciencias matemáticas y la hacen particularmente relevante para otros campos de estudio.

Resulta fundamental que las personas que se dedican a la estadística —o cualquier persona que utilice datos— sean más que simples analistas de datos. Deben resolver problemas con datos, examinarlos en detalle y plantear preguntas a lo largo del proceso de resolución de problemas estadísticos para tomar decisiones con confianza, entendiendo que el arte de comunicar con datos es esencial.

Este informe, por tanto, mejora y actualiza el informe GAISE I de los años 2005 y 2007 para ajustarse a la notable evolución que ha tenido lugar en el ámbito de la estadística en los últimos 15 años. Estas mejoras hacen hincapié en:

- El cuestionamiento a lo largo del proceso de resolución de problemas estadísticos.
- Diferentes datos y tipos de variables.
- El razonamiento multivariante en los niveles A, B y C.
- El razonamiento probabilístico en los niveles A, B y C.
- El papel de la tecnología en la estadística y cómo se desarrolla en cada uno de los niveles.
- Criterios de evaluación que midan el razonamiento estadístico.

Las preguntas en estadística

Independientemente del tipo de datos con los que se esté trabajando (ya sean pequeñas muestras de una población, datos experimentales, etc.), resulta esencial abordar cada uno de los componentes del proceso.

El proceso de resolución de problemas estadísticos empieza típicamente con una pregunta de investigación, seguida de un estudio diseñado para recoger datos que se ajusten a las respuestas a la pregunta. El análisis de los datos también se guía por la pregunta. El cuestionamiento y examen constante de los datos a través del proceso de resolución de problemas estadísticos puede llevar al planteamiento de nuevas preguntas de investigación estadística.

A menudo, cuando se consideran datos secundarios, es necesario examinarlos cuidadosamente primero: cómo se hicieron las mediciones, qué tipo de datos se seleccionaron, cuál es su significado y qué estudio se diseñó para recogerlos. Cuando se dispone de una comprensión más profunda de los datos, se está en disposición de valorar si el conjunto de datos es apropiado para analizar la pregunta de investigación estadística original o si se pueden plantear otras preguntas de investigación estadística para analizarlas con el conjunto de datos secundarios.

El GAISE II modela el uso del cuestionamiento en estadística en todos sus ejemplos. El papel del cuestionamiento y los distintos tipos de preguntas estadísticas empleadas en el proceso de resolución de problemas estadísticos se expone con detalle en Arnold y Franklin (2020).

Diferentes datos y tipos de variables

Las variables tradicionales se clasifican como cualitativas o cuantitativas (numéricas). A su vez, las variables cualitativas se dividen en nominales u ordinales (*ranking*); y las variables cuantitativas, en discretas o continuas, según la escala en que se midan. Los recuentos, como el número de mascotas que tiene un estudiante, son ejemplos de variables cuantitativas discretas; y las mediciones, como la longitud de la cola de una lagartija, son ejemplos de variables cuantitativas continuas. Es esencial que el alumnado se sienta cómodo a la hora de analizar estos tipos de variables tradicionales, y que aprenda a explorar y a describir las características de la distribución de los datos para las variables cualitativas y cuantitativas.

En el mundo actual, las variables también pueden ser imágenes, sonidos, vídeos o palabras. El alumnado debe ser capaz de identificar los datos en bruto de estos tipos de variables no tradicionales, comprender cómo las transformaciones de las variables pueden producir diferentes representaciones de los mismos datos, y organizar estos datos de manera adecuada. Asimismo, tiene que preguntarse cómo se producen los datos —si son primarios (datos recogidos de primera mano) o secundarios (datos que están disponibles).

Cada vez se recogen más datos y proliferan las fuentes de datos secundarios. A menudo, estos datos están disponibles, pero no se encuentran aún listos para el análisis, por lo que el alumnado actual ha de adquirir destrezas para manipularlos y restructurarlos, transformar las variables dadas en nuevas variables y cuestionarse los orígenes y la idoneidad de los datos para el propósito que se persigue (p. ej., preguntándose si una publicación en una red social puede generalizarse a toda la población). Los ejemplos que se exponen en el presente informe ilustrarán cómo trabajar con diferentes tipos de variables.

Pensamiento multivariante en estadística

El pensamiento multivariante debe comenzar en edades tempranas, ya que es natural que el alumnado pregunte y razone con más de dos variables a la vez, especialmente al explorar asociaciones entre variables. Las personas más jóvenes pueden darse cuenta de las diferentes características de una unidad observacional —por ejemplo, ellas mismas y sus compañeros y compañeras de clase—, y registrar esos datos. Hay que animar a todo el alumnado, incluso al más joven, a establecer comparaciones entre grupos. A medida que cada cual avanza en su aprendizaje estadístico, desarrolla las herramientas estadísticas para formalizar comparaciones y asociaciones.

La probabilidad en estadística

La probabilidad tiene que ver con cuantificar la aleatoriedad. Es la base de cómo hacemos las predicciones en estadística. El alumnado, desde una edad temprana, puede utilizar la probabilidad de manera informal para predecir cuán probable o improbable puede ser un suceso en particular, así como considerar predicciones informales más allá del alcance de los datos analizados.

La probabilidad es una herramienta esencial en estadística. La probabilidad también es importante en matemáticas, dado que en esta disciplina se utilizan enfoques y razonamientos diferentes a la estadística. Los dos problemas siguientes y la naturaleza de sus soluciones ilustrarán la diferencia.

Problema 1: Supongamos que un dado es «justo».
Pregunta: Si lanzamos un dado diez veces, ¿cuántas veces obtendremos un número par?

Problema 2: Toma un dado.
Pregunta: ¿Este dado es «justo»? Es decir, ¿cada una de sus caras tiene la misma probabilidad de aparecer?

El problema 1 es un problema de probabilidad matemática. El problema 2 es un problema estadístico que puede usar el modelo matemático de probabilidad determinado en el problema 1 como una herramienta para buscar la solución.

La respuesta no es determinista para ninguna de las dos preguntas. El lanzamiento de dados produce resultados aleatorios, lo que sugiere que la respuesta es probabilística. La solución al problema 1 parte del supuesto de que el dado es justo, y consiste en deducir lógicamente los valores de las probabilidades de cada posible recuento de números pares resultantes de 10 tiradas. Los casos posibles son 0, 1, …, 10.

La solución al problema 2 parte de un dado desconocido que no sabemos si es justo o sesgado. La búsqueda de la respuesta es experimental: tiramos el dado, vemos lo que sucede y examinamos los datos resultantes para ver si parecen provenir de un dado justo o de uno sesgado. Un posible enfoque para emitir este juicio podría ser el siguiente: tira el dado 10 veces y registra cuántos números pares aparecen. Repite este proceso de lanzar el dado 100 veces. Recopila la frecuencia de números pares para cada una de esas 100 repeticiones (p. ej., 5, 3, 6…). Compara estos resultados con las frecuencias pronosticadas por el modelo matemático para un dado justo en el problema 1. Si las frecuencias empíricas del experimento son bastante diferentes de las predichas por el modelo matemático para un dado justo y no es probable que las frecuencias observadas se deban a la variabilidad del azar, entonces podemos concluir que el dado no es justo.

En el problema 1 conformamos nuestra respuesta a partir de deducciones lógicas, y en el problema 2 la conformamos observando resultados experimentales. Al respecto, Gelman y Nolan (2002) ofrecen ideas similares sobre el estudio de nociones básicas de probabilidad.

A menudo, los cálculos probabilísticos pueden aproximarse a través de la simulación, tanto a mano como usando la tecnología. Las simulaciones ayudan al alumnado a alcanzar una comprensión conceptual de cálculos probabilísticos complejos sin depender del cálculo matemático. El nivel C contiene una discusión más profunda sobre el papel de la probabilidad en el razonamiento estadístico.

La probabilidad se usa también en estadística a través de la aleatoriedad —muestras y asignaciones aleatorias—. Las muestras se pueden recoger al azar, y los experimentos se pueden diseñar por medio de asignaciones aleatorias de individuos a diferentes tratamientos. La aleatorización minimiza los sesgos en las selecciones y las asignaciones, y también da lugar a una aleatoriedad en los resultados que puede describirse con modelos de probabilidad.

La tecnología en estadística

La enseñanza de la estadística ha mejorado considerablemente: se ha pasado de enseñar sin utilizar la tecnología a enseñar con la tecnología integrada. Este campo ha evolucionado del uso de lenguajes de programación en los años 80 a calculadoras estadísticas portátiles en la década de los 90 y a calculadoras estadísticas en línea, potentes paquetes de *software* estadístico y herramientas increíbles para la visualización de datos. Hoy en día, la simulación es tan fácil como acceder a un *applet* público donde se da la posibilidad de llevar a cabo miles de repeticiones con solo hacer clic. Los laboratorios informáticos ya no son necesarios, basta con una conexión a internet. El paso a la tecnología web permite un mayor acceso a la visualización, la exploración y la simulación de los datos. Sin embargo, el acceso a la tecnología varía de unas escuelas a otras. No todas las aulas están equipadas con acceso a internet o *hardware* y *software* tecnológicos. Las prácticas estadísticas modernas están entretejidas con tecnología; por lo tanto, se recomienda utilizar la tecnología en la mayor medida posible en cada circunstancia.

El GAISE II muestra cómo incorporar el uso apropiado de la tecnología a las actividades estadísticas en todos los niveles de enseñanza Pre-K–12. En el nivel C se incluye, además, una aproximación más detallada sobre cómo los profesionales de la estadística usan la tecnología.

La evaluación en estadística

Con independencia del tipo de evaluación que se use, los criterios de evaluación deben medir la comprensión conceptual y requerir que el alumnado haga uso del razonamiento estadístico en contexto y de la variabilidad en todas las etapas del proceso de resolución de problemas estadísticos. El GAISE II proporciona ejemplos de proyectos nacionales e internacionales que modelan este tipo de evaluación.

Las evaluaciones deben alinear la tecnología usada para la computación estadística con la tecnología utilizada para la docencia. Por ejemplo, si se utiliza un determinado *software* estadístico para la enseñanza, tiene que utilizarse el mismo *software* en las evaluaciones para, de este modo, evitar que exista un desequilibrio, como sucedería si, por ejemplo, solo se permitiera usar la calculadora en el examen. Más aún, la tecnología debe tenerse en cuenta a la hora de administrar esas evaluaciones, tanto si son formativas como sumativas.

El futuro con los datos

Hoy en día, el arte y la ciencia de trabajar con datos se presentan bajo muchos nombres, entre los que se incluyen la estadística, la ciencia de datos, la informática y el análisis de datos. Todos ellos combinan las habilidades estadísticas, matemáticas y de la ciencia de la computación. Debido a la abundancia de datos que se recogen a diario a través de internet y otros medios de comunicación, el aprendizaje automático, el aprendizaje profundo y la inteligencia artificial son áreas en crecimiento en las que se utilizan grandes conjuntos de datos y se desarrollan algoritmos para realizar predicciones. Por ejemplo, en *marketing* se utilizan algoritmos para predecir el comportamiento de los consumidores. Con estas áreas de estudio, resulta crucial que el proceso de resolución de problemas estadísticos se use constantemente para examinar los datos con detalle. Sin este examen, surgirían sesgos y usos incorrectos. Un ejemplo de ello es el artículo de Angwin et al. (2106), en el que se revelan desigualdades en justicia.

Los datos, tanto en el sector público como en el privado, nos rodean y moldean nuestras vidas profesional y personalmente. Los datos son un medio de comunicación, creación de comunidad y descubrimiento. Los datos se usan para contar una historia. Las personas que se dedican a la estadística ven el mundo a través de los datos —los datos proporcionan modelos de realidad—. El pensamiento estadístico y el proceso de resolución de problemas estadísticos son fundamentales para explorar todos los datos.

EL MARCO DE TRABAJO

La estructura conceptual para la educación estadística se presenta en el modelo basado en el marco de trabajo bidimensional que se muestra en la tabla 1. Una de las dimensiones es definida por los componentes del proceso de resolución de problemas estadísticos que pueden usarse para promover la alfabetización estadística; y la otra dimensión está compuesta por tres niveles de desarrollo.

El proceso de resolución de problemas estadísticos

El propósito del proceso de resolución de problemas estadísticos (véase la figura 3) es recoger y analizar datos para responder preguntas de investigación estadística.

Este proceso de investigación abarca cuatro componentes, cada uno de los cuales conlleva explorar y abordar la variabilidad:

I. Formular preguntas de investigación estadística.
II. Recoger/considerar los datos.
III. Analizar los datos.
IV. Interpretar los resultados.

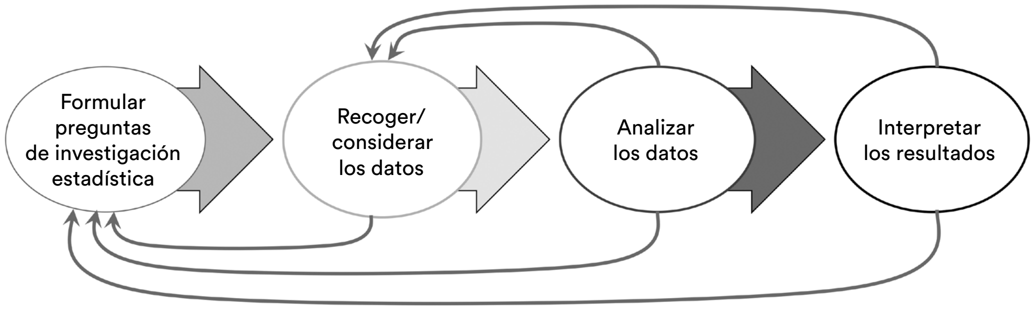


Figura 3. El proceso de resolución de problemas estadísticos

I. Formular preguntas de investigación estadística

Anticipar la variabilidad – Iniciar el proceso

Formular preguntas de investigación estadística que anticipan variabilidad conduce a investigaciones productivas. Por ejemplo, las siguientes preguntas de investigación estadística anticipan variabilidad y podrían conducir a un valioso proceso de recogida de datos y su análisis posterior:

- *¿A qué velocidad crecerá mi planta?*
- *¿Crecen más rápido las plantas expuestas a más luz solar?*
- *¿Cómo afecta la luz solar al crecimiento de una planta?*

Por el contrario, la pregunta «¿Cuánto mide la planta?» se responde con una única altura; por tanto, no es una pregunta de investigación estadística. «¿Cuánto mide la planta?» es una pregunta que hacemos para recoger datos.[6] Se pueden plantear muchas otras preguntas de recogida de datos para ayudar a recopilar los datos necesarios para responder la pregunta de investigación estadística: «¿Crecen más rápido las plantas expuestas a más luz solar?». El hecho de que vaya a haber alturas diferentes para distintas exposiciones a la luz solar implica que anticipamos una respuesta basada en medidas de alturas de plantas que varían.

Las preguntas de investigación estadística marcan el inicio de estudios valiosos, pero el uso del cuestionamiento es prominente a lo largo de los cuatro componentes del proceso de resolución de problemas estadísticos. Estos usos del cuestionamiento se ilustrarán a través de los ejemplos en los diferentes niveles.

Además de la anticipación de variabilidad, hay otras características de una pregunta de investigación estadística que son importantes. La(s) variable(s) de interés, el grupo o población en que se centra la pregunta, y la intención deben ser claros. ¿La pregunta requiere una descripción de los datos?, ¿compara una variable entre dos o más grupos?, ¿busca una asociación entre dos variables? La pregunta debe referirse al grupo completo (anticipando la variabilidad) y no a un individuo (dando una respuesta determinista), debe responderse a través de la recogida de datos (datos primarios) o con los datos disponibles (datos secundarios), y debe tener un propósito.

II. Recoger/considerar los datos

Reconocer la variabilidad – Diseñar teniendo en cuenta las diferencias

Los diseños de recogida de datos deben reconocer la variabilidad en los datos. Algunos métodos de estudio se utilizan para reducir y detectar la variabilidad en los datos, como el control estadístico de procesos y el muestreo aleatorio. Otros se usan para inducir variabilidad y probar tratamientos, como el diseño de experimentos. En este último, se opta por los diseños experimentales para reconocer las diferencias entre grupos sujetos a diferentes tratamientos. La asignación aleatoria a los grupos pretende reducir las diferencias entre grupos debidas a factores no manipulados o controlados en el experimento. En todos los diseños, un aspecto clave del análisis estadístico es identificar, tener en cuenta y explicar la variabilidad.

Una vez que los datos estén disponibles (ya sean recogidos de primera mano, ya sean adquiridos de otra fuente), deben examinarse en detalle. Por ejemplo, es necesario plantearse preguntas como: ¿qué tipo de variables hay?, ¿cuáles son los posibles valores que pueden tomar? y ¿cómo se recogieron los datos? Estas cuestiones ayudan a determinar si los datos son adecuados para responder la pregunta de investigación estadística. El diseño de recogida de datos influye en el alcance de la generalización y en las posibles limitaciones del análisis y la interpretación.

III. Analizar los datos

Considerar la variabilidad – Usar distribuciones

Cuando analizamos datos, buscamos entender su variabilidad. Razonar sobre las distribuciones es clave para considerar y describir la variabilidad en todos los niveles de desarrollo. Las representaciones gráficas y los resúmenes numéricos se usan para explorar, describir y comparar la variabilidad en las distribuciones.

Por ejemplo, los bateos promedio de los equipos de béisbol de la American League y los bateos promedio de los equipos de la National League en un año concreto pueden representarse en gráficos o diagramas de puntos[7] y en diagramas de cajas[8] comparativos. Estos gráficos muestran la variabilidad de la distribución de bateos promedios de ambas ligas. Podemos tener en cuenta la variabilidad mediante la descripción del solapamiento y la separación de las distribuciones de las dos ligas.

[6] N. de T.: La literatura estadística en español se refiere, en ocasiones, a este tipo de preguntas como «preguntas de encuesta». En este fragmento, hemos optado por utilizar «preguntas para recoger datos» a fin de mantener una mayor fidelidad al texto original. Posteriormente, utilizaremos también «preguntas de encuesta» en otros contextos.

[7] N. de T.: Los gráficos de puntos en la literatura en español también aparecen con frecuencia con el término original en inglés, *dotplot*.

[8] N. de T.: Análogamente, los diagramas de caja, también llamados *diagramas de caja y bigotes*, también aparecen frecuentemente con la denominación inglesa, *boxplot*.

Otro ejemplo de la consideración de la variabilidad es el margen de error en las encuestas de opinión pública. Cuando los resultados de una encuesta electoral indican que «el 42 % de las personas encuestadas apoyan a un candidato en particular, con un margen de error de ± 3 puntos porcentuales al nivel de confianza del 95 %», el objetivo del margen de error es tener en cuenta la variabilidad muestral.

IV. Interpretar los resultados

Permitir la variabilidad – Mirar más allá de los datos

Las interpretaciones estadísticas se hacen en presencia de variabilidad y deben tenerla en cuenta. Por ejemplo, tenemos que interpretar los resultados de una encuesta electoral como una estimación que puede variar de una muestra de votantes a otra. Al interpretar los resultados de un ensayo clínico aleatorizado, debemos recordar que existen dos fuentes importantes de variabilidad: la inclusión aleatoria en el grupo de tratamiento y la variabilidad de individuo a individuo. Cuando generalizamos los resultados y miramos más allá de los datos recogidos para el estudio, debemos tener en cuenta estas fuentes de variabilidad.

Tres niveles de desarrollo: A, B y C

Las personas expertas en estadística entienden el papel de la variabilidad en el proceso de resolución de problemas estadísticos. Al formular su primera pregunta, anticipan la recogida de datos, la naturaleza del análisis y las posibles interpretaciones, lo que implica posibles fuentes de variabilidad. Al final, los profesionales experimentados reflexionan sobre todos los aspectos de la recogida y el análisis de datos, así como sobre la propia pregunta al interpretar los resultados. Del mismo modo, vinculan la recogida y el análisis de datos entre sí, así como con los demás componentes del proceso de resolución de problemas estadísticos.

No se puede esperar que el alumnado principiante establezca todas esas conexiones, puesto que se necesitan años de experiencia y entrenamiento para desarrollar un razonamiento más maduro. Al igual que la educación matemática, la educación estadística debe verse como un proceso de desarrollo.

Para alcanzar los objetivos de la alfabetización estadística, este informe, como el GAISE I, proporciona un marco de trabajo distribuido en tres niveles —A, B y C— para la educación estadística en el contexto de Pre-K–12. Los estudiantes tienen nociones innatas de variabilidad y probabilidad a edades muy tempranas. En Leavy et al. (2018) se resume la investigación existente al respecto. El nivel A, partiendo de este conocimiento innato, presenta el proceso de resolución de problemas estadísticos de manera más formal al alumnado. El nivel B prosigue con la construcción del repertorio de herramientas estadísticas. Y en el nivel C ya se le pueden proporcionar objetivos de aprendizaje más elevados para el desarrollo de la alfabetización estadística en la educación Pre-K–12. Por ello, el nivel C establece objetivos ambiciosos para el alumnado que finaliza la etapa Pre-K–12 en la sociedad actual, impulsada por los datos. De este modo, se le prepara para que vaya más allá de estos niveles y desarrolle preguntas de investigación estadística y técnicas de análisis más complejas, mientras trabaja con tipos de datos en constante evolución.

A pesar de que estos tres niveles pueden equipararse a cursos académicos, se basan en el desarrollo en la alfabetización estadística, no en la edad. No se intenta vincular *de ninguna forma* estos niveles a cursos específicos. Así, una estudiante de secundaria que no tenga ninguna experiencia previa con la estadística deberá comenzar con los conceptos y las actividades del nivel A antes de progresar al nivel B. Este prerrequisito se aplica también al alumnado de bachillerato. Si un estudiante no ha tenido experiencias de los niveles A y B antes de bachillerato, entonces no es apropiado que empiece asumiendo las expectativas del nivel C. En el nivel A, las investigaciones y las situaciones hipotéticas están más dirigidas por el profesorado, mientras que, en los niveles B y C, están más dirigidas por el alumnado.

La tabla del marco de trabajo

Al igual que en la tabla del marco de trabajo del GAISE I, se describe aquí el desarrollo de cada una de las cuatro etapas o componentes del proceso en todos los niveles. Se entiende que el trabajo del nivel B asume y continúa desarrollando los conceptos del nivel A; asimismo, el nivel C asume y utiliza conceptos de los niveles más bajos. Los aspectos esenciales del GAISE I son similares en el GAISE II, pero mejorados con más detalles y algunos elementos adicionales para tener en cuenta la evolución de la estadística desde el GAISE I.

Cada columna describe un problema de investigación completo para un nivel específico.

Tabla 1. El marco de trabajo

Componente del proceso	Nivel A	Nivel B	Nivel C
I. Formular preguntas de investigación estadística	Comprender cuándo una investigación estadística es apropiada. Plantear preguntas de investigación estadística de interés para el alumnado en las que el contexto le permita recoger todos los datos necesarios o acceder a ellos. Plantear preguntas de investigación estadística de resumen (o descriptivas) sobre una variable respecto a grupos pequeños y bien definidos (p. ej., un subconjunto de una clase, la clase, la escuela, la ciudad) y ampliarlas para incluir preguntas de investigación estadística de comparación y asociación entre variables. Experimentar con diferentes tipos de preguntas en estadística: las que se usan para enmarcar una investigación, las que se usan para recoger datos y las que se usan para guiar el análisis y la interpretación.	Reconocer que las preguntas de investigación estadística pueden usarse para definir temas de investigación, y que se pueden plantear muchas preguntas de investigación estadística sobre cualquier tema. Entender que las preguntas de investigación estadística tienen en cuenta el contexto y la variabilidad presente en los datos. Plantear preguntas de investigación estadística de resumen, comparativas y de asociación sobre una población más amplia utilizando muestras extraídas de esa población. Plantear preguntas de investigación estadística que requieran observar una variable a lo largo del tiempo. Entender que existen diferentes tipos de preguntas en estadística: las que se usan para enmarcar una investigación, las que se usan para recoger datos y las que se usan para guiar el análisis y la interpretación. Plantear preguntas de investigación estadística a partir de datos recogidos de fuentes en línea y páginas web, *smartphones*, dispositivos de actividad física, sensores y otros dispositivos actuales.	Formular preguntas de investigación estadística multivariante y determinar de qué manera se pueden recoger y analizar los datos para dar una respuesta. Plantear preguntas de investigación estadística de resumen, comparación y asociación para encuestas, estudios observacionales y experimentos que usen datos primarios o secundarios. Plantear preguntas de investigación de inferencia estadística sobre la causalidad y la predicción.

Componente del proceso	Nivel A	Nivel B	Nivel C
II. Recoger/ considerar los datos	Entender que los datos son información; asumir que, para responder una pregunta de investigación estadística, una persona puede recoger datos específicamente para ese fin o usar datos que hayan sido recogidos por otras personas con fines diferentes. Entender cómo recoger y registrar información de un grupo de interés mediante encuestas y mediciones recogidas a partir de observaciones y experimentos simples. Entender que una variable mide la misma característica en varios individuos u objetos, lo que implica que los valores de los datos pueden fluctuar. Entender que en un conjunto de datos puede haber diferentes tipos de variables (p. ej., cualitativas o cuantitativas). Examinar en detalle el conjunto de datos para entender el contexto de las variables, ya que pueden estar relacionadas con preguntas de investigación estadística. Entender que los datos no siempre están limpios, sino que pueden contener errores, tener valores perdidos, etc., y que deben tomarse decisiones sobre cómo abordar estos problemas.	Entender que los datos son información recogida y registrada con un propósito y que se pueden organizar y almacenar en distintas estructuras (p. ej., hojas de cálculo). Comprender que una muestra puede usarse para responder preguntas de investigación estadística sobre una población. Reconocer las limitaciones y el alcance de los datos recogidos mediante la descripción del grupo, o población, del que se han obtenido. Entender que los datos pueden usarse para hacer comparaciones entre diferentes grupos en un único momento temporal o dentro del mismo grupo a lo largo del tiempo. Reconocer que los datos pueden ser recogidos utilizando encuestas y mediciones, y desarrollar una actitud crítica al analizar los métodos de recogida de datos. Entender que las variables cuantitativas pueden ser discretas o continuas. Entender cómo examinar los datos en detalle para determinar de qué manera fueron recogidos, de quién se obtuvieron, qué tipos de variables contienen, cómo se midieron esas variables (incluyendo las unidades usadas) y sus posibles resultados. Comprender que se pueden recoger datos (datos primarios) u obtener datos ya existentes de otras fuentes (datos secundarios). Entender cómo se usa la asignación aleatoria en experimentos comparativos para controlar las características que podrían afectar a las respuestas.	Aplicar un plan adecuado de recogida de datos primarios o selección de datos secundarios para la pregunta de investigación estadística de interés. Distinguir entre encuestas, estudios observacionales y experimentos. Entender qué constituye una buena práctica para diseñar un estudio muestral, un experimento y un estudio observacional. Entender el papel de la selección aleatoria en los estudios muestrales y el efecto del tamaño muestral en la variabilidad de las estimaciones. Entender el papel de la asignación aleatoria en los experimentos y sus implicaciones en las interpretaciones causa-efecto. Entender los problemas de los sesgos y de las variables de confusión en los estudios observacionales, y sus implicaciones para la interpretación. Entender las prácticas de manejo de datos que mejoran la reproducibilidad y aseguran un uso ético, incluidas las descripciones de las alteraciones, y comprender cuándo los datos pueden contener información sensible. Entender cómo la preocupación por la privacidad y por los seres humanos puede afectar a la recogida y la distribución de los datos. Entender que, en algunas circunstancias, los datos recogidos o considerados podrían no ser generalizables a la población objetivo, o que estos datos pueden ser toda la población.

Componente del proceso	Nivel A	Nivel B	Nivel C
III. Analizar los datos	Entender que la distribución de una variable cualitativa o cuantitativa describe el número de veces que se da un resultado concreto. Representar la variabilidad de variables cualitativas o cuantitativas utilizando las representaciones adecuadas (p. ej., tablas, gráficos de recuento icónico, gráficos de puntos, gráficos de barras). Describir las características clave de las distribuciones de variables cuantitativas, como: ○ centrales (la media como el reparto equitativo, y la mediana como el valor central de los datos ordenados); ○ variabilidad (el rango como la diferencia entre el mayor y el menor valor, y la dispersión como las unidades de distancia al valor del reparto equitativo); ○ forma (agrupaciones, simétricas o no, y espacios vacíos). Reconocer que las distribuciones pueden usarse para comparar dos grupos. Observar si parece haber una asociación entre dos variables.	Representar la variabilidad de variables utilizando las representaciones adecuadas (p. ej., gráficos de puntos, gráficos de barras). Aprender a utilizar las características clave de las distribuciones de variables cuantitativas, como: ○ centrales (la media como el punto de equilibrio, y la mediana como el valor central de los datos ordenados); ○ variabilidad (el rango intercuartílico y la desviación absoluta media [DAM]); ○ forma (simétrica o asimétrica, y número de modas). Usar el razonamiento sobre las distribuciones para comparar dos grupos a partir de variables cuantitativas. Explorar patrones de asociación entre dos variables cuantitativas o cualitativas: ○ medidas de correlación (razón de recuento de cuadrantes [RRC]); ○ comparación de proporciones condicionadas entre variables cualitativas.	Usar la tecnología para dividir y filtrar conjuntos de datos y transformar variables, incluyendo el suavizado de series temporales. Identificar formas adecuadas de resumir datos cuantitativos o cualitativos usando tablas, representaciones gráficas y resúmenes numéricos estadísticos, lo que incluye el uso de la desviación típica como medida de variabilidad, y el diagrama de caja modificado para identificar valores atípicos. Resumir y describir relaciones entre múltiples variables. Entender cómo se usan las distribuciones muestrales (desarrolladas mediante simulación) para describir la variabilidad entre muestras de los estadísticos muestrales. Desarrollar simulaciones para determinar distribuciones muestrales aproximadas y calcular p-valores a partir de esas distribuciones. Describir la asociación entre dos variables cualitativas, usando medidas como la diferencia de proporciones o el riesgo relativo. Describir la relación entre dos variables cuantitativas mediante la interpretación del coeficiente de correlación de Pearson y de la recta de regresión de mínimos cuadrados. Usar simulaciones para investigar las asociaciones entre dos variables cualitativas y para comparar grupos. Construir intervalos de predicción e intervalos de confianza a fin de determinar valores posibles para una observación predicha o una característica de la población.

Componente del proceso	Nivel A	Nivel B	Nivel C
IV. Interpretar los resultados	Usar la evidencia estadística obtenida en los análisis para responder las preguntas de investigación estadística y comunicar los resultados mediante respuestas estructuradas con la orientación del profesorado. Hacer afirmaciones sobre el grupo, o población, del que se recogieron los datos, reconociendo que las conclusiones se circunscriben a estos grupos y no pueden generalizarse a otros. Describir las diferencias entre dos grupos bajo condiciones distintas.	Utilizar la evidencia estadística obtenida en los análisis para responder las preguntas de investigación estadística y comunicar los resultados mediante respuestas completas, con cierta orientación del profesorado. Reconocer que los datos se pueden leer de una manera más profunda. Generalizar a partir de la muestra proporcionando evidencia estadística para la generalización y mencionando la incertidumbre y la verosimilitud cuando sea necesario. Reconocer la incertidumbre causada por la variabilidad entre las muestras. Exponer las limitaciones de la información muestral (p. ej., una muestra puede ser o no representativa de una población más grande, variabilidad de la medición). Comparar los resultados de un experimento obtenidos bajo condiciones diferentes.	Utilizar la evidencia estadística obtenida en los análisis para responder las preguntas de investigación estadística y comunicar los resultados por medio de informes y presentaciones más formales. Evaluar e interpretar el impacto de los valores atípicos en los resultados. Entender qué significa que un resultado o una estimación de una característica de la población sea posible o no en comparación con la variación debida al azar. Interpretar el margen de error asociado a una estimación de una característica de la población. Reconocer la existencia de valores perdidos y entender cómo esos valores pueden añadir sesgos al análisis. Usar el pensamiento multivariante para entender cómo influyen unas variables en otras. Comunicar a otras personas el razonamiento estadístico y sus resultados en una variedad de formatos (verbal, escrito, visual). Entender cómo interpretar adecuadamente los p-valores simulados.

Nivel A

Introducción

El alumnado está rodeado de datos, y es posible que los conciba como el recuento de sus preferencias —por ejemplo, el tipo de música favorito— o como mediciones —por ejemplo, la longitud de los brazos o el número de libros que hay en la mochila. El alumnado del nivel A puede estar en educación primaria, secundaria o, incluso, en cursos superiores; o ser personas adultas que no están escolarizadas. Independientemente de la edad, todo el mundo debería empezar sus estudios de estadística en este nivel.

En el nivel A se desarrolla un sentido de los datos; es decir, se comprende que los datos son información. Por otra parte, se aprende que los datos se generan en contextos o situaciones específicos, y que se pueden usar para responder preguntas de investigación estadística sobre un determinado contexto o situación. También se empieza a aprender cómo examinar los datos en detalle.

El alumnado debe tener oportunidades de generar preguntas de investigación estadística sobre un contexto específico (p. ej., su clase) y determinar qué datos se tienen que recoger o recuperar para responder esas preguntas.

Asimismo, debe aprender cómo representar gráficamente sus datos, describir características de las distribuciones y empezar a usar estas descripciones para responder las preguntas de investigación estadística planteadas.

Finalmente, el alumnado de este nivel se debe formar unas ideas iniciales sobre cómo se relacionan la probabilidad y el razonamiento estadístico. Estas ideas le servirán al utilizar la probabilidad para hacer inferencias, informalmente, en el nivel B y, más formalmente, en el nivel C.

En el nivel A, el alumnado puede recoger los datos, o bien se le pueden proporcionar datos secundarios. El profesorado debe aprovechar las situaciones que se produzcan de manera natural y en las que el alumnado perciba un patrón respecto a algunos datos y comience a plantearse preguntas. Por ejemplo, si un día, al pasar lista por la mañana, alguien se da cuenta de que hay muchas ausencias, la profesora puede aprovechar la oportunidad para hacer que la clase formule preguntas de investigación estadística que podrían responderse con los datos de asistencia.

Aspectos esenciales de cada componente

Las recomendaciones del nivel A en el proceso de resolución de problemas estadísticos son:

I. Formular preguntas de investigación estadística

- → Comprender cuándo una investigación estadística es apropiada.
- → Plantear preguntas de investigación estadística de interés para el alumnado en las que el contexto le permita recoger todos los datos necesarios o acceder a ellos.
- → Plantear preguntas de investigación estadística de resumen (o descriptivas) sobre una variable respecto a grupos pequeños y bien definidos (p. ej., un subconjunto de una clase, la clase, la escuela, la ciudad) y ampliarlas para incluir preguntas de investigación estadística de comparación y asociación entre variables.
- → Experimentar con diferentes tipos de preguntas en estadística: las que se usan para enmarcar una investigación, las que se usan para recoger datos y las que se usan para guiar el análisis y la interpretación.

II. Recoger/considerar los datos

- → Entender que los datos son información; asumir que, para responder una pregunta de investigación estadística, una persona puede recoger datos específicamente para ese fin o usar datos que hayan sido recogidos por otras personas con fines diferentes.
- → Entender cómo recoger y registrar información de un grupo de interés mediante encuestas y mediciones recogidas a partir de observaciones y experimentos simples.
- → Entender que una variable mide la misma característica en varios individuos u objetos, lo que implica que los valores de los datos pueden fluctuar.
- → Entender que en un conjunto de datos puede haber diferentes tipos de variables (p. ej., cualitativas o cuantitativas).
- → Examinar en detalle el conjunto de datos para entender el contexto de las variables, ya que pueden estar relacionadas con preguntas de investigación estadística.
- → Entender que los datos no siempre están limpios, sino que pueden contener errores, tener valores perdidos, etc., y que deben tomarse decisiones sobre cómo abordar estos problemas.

III. Analizar los datos

- → Entender que la distribución de una variable cualitativa o cuantitativa describe el número de veces que se da un resultado concreto.
- → Representar la variabilidad de variables cualitativas o cuantitativas utilizando las representaciones adecuadas (p. ej., tablas,[9] gráficos de recuento icónico,[10] gráficos de puntos, gráficos de barras).
- → Describir las características clave de las distribuciones de variables cuantitativas, como:
  - o centrales (la media como el reparto equitativo, y la mediana como el valor central de los datos ordenados);
  - o variabilidad (el rango como la diferencia entre el mayor y el menor valor, y la dispersión como las unidades de distancia al valor del reparto equitativo);
  - o forma (agrupaciones, simétricas o no, y espacios vacíos).
- → Reconocer que las distribuciones pueden usarse para comparar dos grupos.
- → Observar si parece haber una asociación entre dos variables.

IV. Interpretar los resultados

- → Usar la evidencia estadística obtenida en los análisis para responder las preguntas de investigación estadística y comunicar los resultados mediante respuestas estructuradas con la orientación del profesorado.
- → Hacer afirmaciones sobre el grupo, o población, del que se recogieron los datos, reconociendo que las conclusiones se circunscriben a estos grupos y no pueden generalizarse a otros.
- → Describir las diferencias entre dos grupos bajo condiciones distintas.

---

[9] N. de T.: En el original, el término *tables* se usa para referirse a las tablas de recogida de datos sin procesar (véase la tabla 2); en las tablas estadísticas de frecuencias, en las que se expresan cuantitativamente las frecuencias de cada categoría de la variable (véase la tabla 3); o bien, más adelante, en las tablas de doble entrada o de contingencia (véase la tabla 10). En el contexto español, el término *tabla* se usa también para referirse a las tablas estadísticas de recuento, las que se utilizan para organizar los datos y facilitar el recuento. Las tablas estadísticas de recuento requieren indicar las categorías de la variable en uno de los ejes, y pueden ser concretas (cuando se usan objetos o imágenes de objetos para hacer el recuento) o pictóricas (cuando se usan signos como cruces, barras... a modo de marcas de recuento).

[10] N. de T.: En el original se diferencian los gráficos de recuento icónico (*picture graphs*), que requieren indicar las categorías de la variable en uno de los ejes y el intervalo frecuencial en el otro (véase la figura 4), de los pictogramas (*pictographs*). En los de recuento icónico, cada símbolo o icono representa una única observación; mientras que, en los pictogramas, cada icono representa una cantidad de observaciones mayor que uno. En el contexto hispanohablante, la definición de *pictograma* es más inclusiva, por lo que los gráficos de recuento icónico aparecen con frecuencia también referidos como *pictogramas*. Aquí se conservan las denominaciones para respetar el material original, ya que, más adelante, el informe diferenciará entre ambos tipos de representación gráfica.

Ejemplo 1: La elección de la música para la fiesta escolar. Realización de un sondeo y resumen de datos

Formular preguntas de investigación estadística

El alumnado del nivel A puede tener interés por el tipo de música favorito de sus compañeros y compañeras. Imaginemos que se está planificando una fiesta escolar para determinado curso y solo hay suficiente dinero para contratar a un único grupo musical. Una clase de ese curso podría plantear la pregunta de investigación estadística:

¿Qué tipo de música le gusta al alumnado de nuestro curso?

Esta pregunta de investigación estadística intenta medir una característica, la preferencia en el tipo de música, de la población de estudiantes de ese curso (nótese que para los más jóvenes o inexpertos tendría sentido investigar la pregunta: «¿Qué tipo de música le gusta al alumnado de nuestra clase?», ya que la pregunta planteada arriba requerirá algunas inferencias desde el nivel de la clase hasta el del curso).

Recoger/considerar los datos

Para responder la pregunta de investigación estadística, el alumnado debe recoger datos sobre la música que le gusta. Antes de comenzar la recogida, sin embargo, es importante reflexionar sobre los métodos de recogida de datos.

Una encuesta es un método de recogida de datos natural para el nivel A. Una posible pregunta de la encuesta podría ser: «¿Cuál es tu tipo de música favorito?». Sin embargo, este formato de pregunta puede dar lugar a muchas respuestas diferentes, lo que podría complicar el análisis de los datos. Tras debatir los pros y los contras de una pregunta abierta o más restringida, se podría modificar la pregunta de la encuesta a: «¿Cuál es tu tipo de música favorito: *country*, rap o *rock*?». Dado que esta pregunta pide específicamente que la persona encuestada escoja entre tres opciones, será más fácil manejar y analizar los datos. El inconveniente de esta pregunta es que restringe las opciones de las personas encuestadas, por lo que, en el caso de alguien que prefiera el *jazz*, su respuesta no indicará su música favorita.

El tipo de música es una variable cualitativa definida aquí por *country*, rap o *rock*. Los datos que se obtienen cuando cada estudiante indica su preferencia en cuanto al tipo de música se denominan *datos cualitativos*.

Cuando el alumnado determina la pregunta de la encuesta, puede llevar a cabo un censo para que cada miembro de la clase responda la pregunta. El alumnado del nivel A debería reconocer que existe variabilidad de un individuo a otro. William Osler, un famoso médico de principios del siglo XX, dijo que la variabilidad es la ley de la vida, y así como no hay dos rostros iguales, tampoco hay dos cuerpos iguales, y ningún par de individuos reacciona o se comporta de la misma manera bajo las condiciones anormales que conocemos como enfermedad (Silverman et al., 2008).

El análisis de los resultados de esta clase se usará para inferir cuál podría ser el tipo de música favorito de todo el curso.

Supongamos que la encuesta realizada a 24 estudiantes de una clase ha generado los datos que se muestran en la tabla 2.

Tabla 2. Datos recogidos, sin procesar

Nombre	Música
Aaron	*Country*
Aden	Rap
Alex	Rap
Angelica	*Rock*
Ana	*Country*
Ariella	*Country*
Eliana	Rap
Elizabeth	Rap

Nombre	Música
Emilio	Rap
Evangeline	*Rock*
Felicity	*Country*
Gabriel	Rap
Isabel	Rap
Jake	Rap
Jerry	*Country*
Leo	*Country*

Nombre	Música
Maria	*Rock*
Micheal	Rap
Nat	*Country*
Penny	Rap
Sofie	Rap
Veronica	*Country*
Vicki	Rap
Xavier	Rap

Analizar los datos

Existen diversas formas de organizar y representar los datos sin procesar. Por ejemplo, un estudiante de poca edad puede crear un gráfico de barras formando filas según su tipo de música favorito, o representar su categoría mediante notas adhesivas (pósits) en la pizarra o en el suelo, para, después, contar el número de estudiantes que hay en cada fila o el número de notas adhesivas que hay en cada categoría.

En el nivel A se puede usar también un gráfico de recuento icónico para representar la distribución de la variable cualitativa «Tipo de música». La distribución resume los datos de esta variable e identifica las frecuencias de cada una de las tres categorías. El gráfico de recuento icónico usa algún tipo de imagen (como un instrumento musical) para representar la preferencia de un individuo. Véase la figura 4 como ejemplo. De este modo, cada estudiante que tenga un tipo de música favorito colocará un recorte con la imagen del instrumento directamente en el gráfico creado por el profesor.

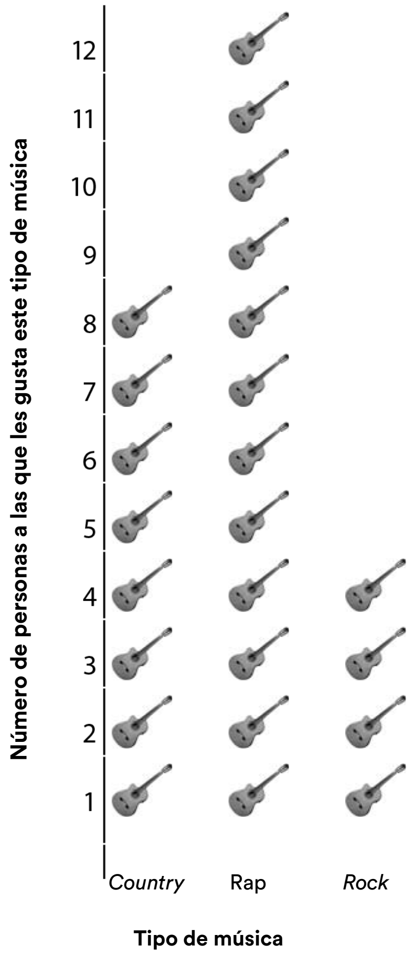


Figura 4. Gráfico de recuento icónico de preferencias musicales

En lugar de una imagen de un instrumento, puede usarse otra representación gráfica (como una X o un cuadrado coloreado) para representar la preferencia de cada persona.

Nótese la diferencia entre un gráfico de recuento icónico y un pictograma. En un gráfico de recuento icónico se utiliza un objeto, como un recorte de cartulina, para representar a un individuo en el gráfico. En cambio, en un pictograma se usa una imagen o un símbolo para representar varios ítems que pertenecen a la misma categoría.

Por ejemplo, en un pictograma que muestre la distribución del alumnado de una clase que van en coche, andando o en autobús, se podría utilizar un recorte de un autobús escolar para representar a cinco pasajeros. Por ejemplo, si en la clase hay 13 personas que van en autobús, habría aproximadamente 2,5 autobuses en el gráfico.

En cualquiera de los dos tipos de gráfico, si se emplean varios símbolos, es importante que estos sean del mismo tamaño y que estén equiespaciados entre ellos para evitar distorsionar visualmente los datos. Por ejemplo, si en la figura 4 se usasen imágenes de una guitarra, un micrófono y un tambor, los dibujos deberían tener la misma altura y anchura.

Otra representación gráfica común para variables cualitativas es un gráfico circular (diagrama de sectores). Estos gráficos pueden ser útiles para las fracciones sencillas, como las mitades y los cuartos. Sin embargo, también requieren la comprensión del razonamiento proporcional y, por lo tanto, deberían usarse solo con el alumnado que haya desarrollado estas habilidades.

Los datos en bruto de la encuesta de preferencias musicales se pueden resumir en la tabla de frecuencias que se presenta en la tabla 3. Esta tabla de frecuencias es una representación tabular que resume los datos cualitativos en bruto. El alumnado puede usar primero marcas de conteo para registrar los datos cualitativos antes de calcular las frecuencias (recuentos) para cada categoría.

Tabla 3. Tabla de frecuencias de preferencias musicales

Favorito	Frecuencia
Country	8
Rap	12
Rock	4

El alumnado del nivel A también debería familiarizarse con los gráficos de barras. Un gráfico de barras resume los datos de alguna otra representación, como un gráfico de recuento icónico o una tabla de frecuencias. La figura 5 muestra el gráfico de barras de las preferencias musicales que habían sido representadas en la tabla de frecuencias y en el gráfico de recuento icónico. Nótese que, como los datos son cualitativos, las categorías del gráfico de barras pueden colocarse en cualquier orden.

En el nivel A se debe aprender que la moda es el resultado más común o más frecuente de una variable. La moda es una medida de resumen útil para los datos cualitativos. Asimismo, se debe comprender que la moda es la categoría que contiene el mayor número de datos, y se la suele denominar *categoría modal*. Si dos categorías están «empatadas» o contienen el mismo número de valores, la distribución es bimodal.

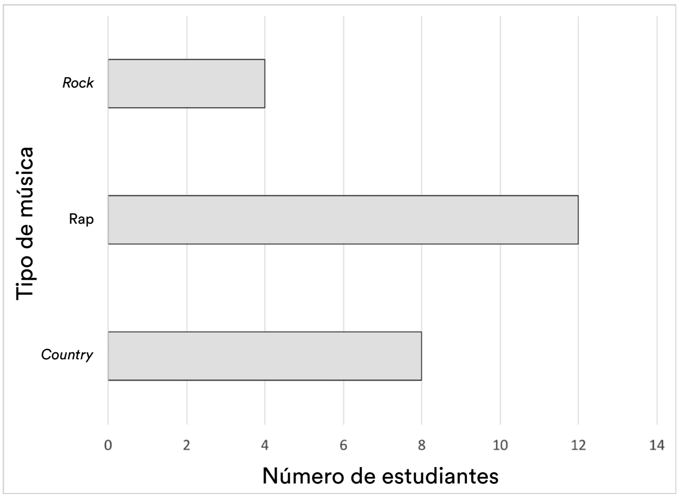


Figura 5. Gráfico de barras del tipo de música favorita de los estudiantes de la clase

Igualmente, se debe reconocer la moda como una manera de describir un valor «representativo» o «típico» de la distribución. En el ejemplo de las preferencias musicales, la música rap fue la preferida por más estudiantes, por delante de cualquier otro tipo de música; por tanto, la moda o categoría modal del conjunto de datos es la música rap. El alumnado ha de ver que la categoría modal (rap) no contiene, necesariamente, la mayoría de las respuestas. En este ejemplo, hay 12 estudiantes que prefieren el rap, pero hay 12 estudiantes que no prefieren el rap.

Interpretar los resultados

En este punto de la investigación, el profesorado debe instar al alumnado a responder la pregunta de investigación estadística inicial:

¿Qué tipo de música le gusta al alumnado de nuestro curso?

Una respuesta potencial de una estudiante podría ser:

«El tipo de música más popular en nuestra clase fue el rap. Por tanto, asumimos o inferimos que el rap será el tipo de música más popular en nuestro curso».

En el nivel A se debe animar al alumnado para que profundice en tal respuesta e incluya una reflexión sobre la variabilidad de los datos. Por ejemplo, la respuesta anterior podría incluir la siguiente descripción:

«Un total de 12 estudiantes prefirieron el rap, mientras que solo 8 prefirieron el *country* y 4 prefirieron el *rock*. Hubo 8 estudiantes más que prefirieron el rap en lugar del *rock*».

Si el alumnado ha desarrollado el razonamiento multiplicativo, también podría decir:

«Esto muestra que hubo tres veces más estudiantes que prefirieron la categoría más popular en comparación con el número de estudiantes que prefirieron la categoría menos popular».

En el nivel A se debe empezar a entender que la probabilidad es una medida de la posibilidad de que algo ocurra. Es una medida del grado de certeza o de incertidumbre. La probabilidad de los sucesos debe verse como un continuo de lo imposible a lo seguro, con niveles intermedios como menos probable, igualmente probable y más probable. Utilizando estas nociones probabilísticas, el alumnado puede profundizar en sus respuestas a la pregunta de investigación estadística para incluir enunciados como:

Si se elige al azar a una persona de nuestra clase, ¿es más probable que prefiera el rap, el country *o el* rock*?*

La mayoría de los datos recogidos en el nivel A requerirán de un censo del alumnado de la clase. Así que el primer paso es que el alumnado lea e interprete, de una manera simple, qué dicen los datos de su clase. Recuérdese que la pregunta de investigación estadística original se refiere a un curso determinado. Debe animarse al alumnado para que reflexione sobre si los resultados obtenidos de los datos de la clase podrían ser representativos de otras clases de su mismo curso y si se podrían extrapolar a este grupo más amplio.

Dado que en el nivel B se recogerán datos de grupos más amplios, en el nivel A se debería promover que el alumnado piense sobre grupos que abarquen más de una clase, para prepararlo de cara al siguiente nivel. Estos grupos más amplios podrían incluir al alumnado de todo el centro, de toda la zona educativa o de toda la región, e, incluso, a toda la población del país. El alumnado debe darse cuenta de qué variables (como la edad o la ubicación geográfica) podrían ser diferentes en el grupo más amplio.

Siguiendo con el ejemplo anterior sobre la música, el alumnado podría plantearse que, si recogiera datos sobre el tipo de música favorito de sus docentes, es probable que el tipo de música resultante fuera distinto. También podría considerar qué ocurriría si recogiera las preferencias musicales del alumnado de bachillerato de su centro educativo.

Ejemplo 2: El tamaño de la familia. La media como reparto equitativo/ justo y la variabilidad como el número de pasos

Formular preguntas de investigación estadística

Supongamos que un instituto está interesado en saber cuántas personas conviven en el hogar de cada estudiante. El alumnado del nivel A podría querer hacer una pregunta similar acerca de cuántas personas conviven en el hogar de sus compañeros y compañeras de clase. El profesorado debe orientar al alumnado hacia el planteamiento de una pregunta de investigación estadística precisa que incluya a la población objetivo (p. ej., los hogares del alumnado de la profesora López) y la variable que va a medirse (p. ej., el número de personas que vive en cada hogar), y que anticipe la variabilidad (p. ej., preguntar por «el número de convivientes» anticipa variabilidad, mientras que preguntar por «el número típico de convivientes» sugiere una respuesta determinista, similar a preguntar cuál es el tamaño promedio de la familia (nótese que esta es una pregunta analítica, que plantearíamos al hacer el análisis). Un ejemplo así podría ser:

¿Cuál es el tamaño de la familia típica del alumnado de la profesora López?

Recoger/considerar los datos

Supongamos que la profesora decide trabajar con nueve estudiantes de su clase, y le pregunta a cada uno: «¿Cuántas personas, incluyéndote a ti, conviven en la casa donde vives la mayor parte del año?». Esta es la pregunta de encuesta que ayuda a responder la pregunta de investigación estadística planteada anteriormente. Cada estudiante representa el tamaño de su familia con una colección de policubos.

Los datos para el «tamaño de la familia» se representan con policubos, como se muestra en la figura 6.

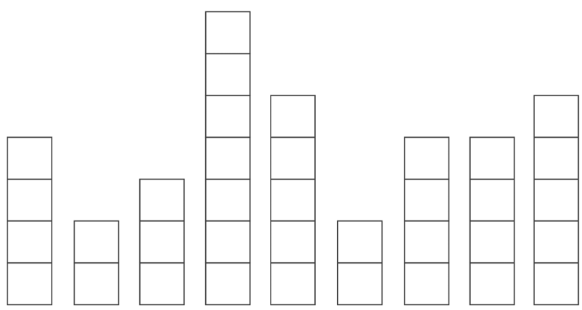


Figura 6. Policubos apilados que representan el tamaño de las familias.

Analizar los datos

Para examinar la distribución del tamaño de las familias para los datos recogidos, el alumnado debe colocar las pilas de policubos en orden creciente (véase la figura 7).

El alumnado debe reconocer que los tamaños de las familias varían. Una pregunta de análisis que la profesora podría plantear es:

¿Cuánta gente habría en cada familia si las nueve fueran del mismo tamaño?

Cuando hacemos que todas las familias tengan el mismo tamaño, este no varía. El alumnado puede usar dos enfoques diferentes:

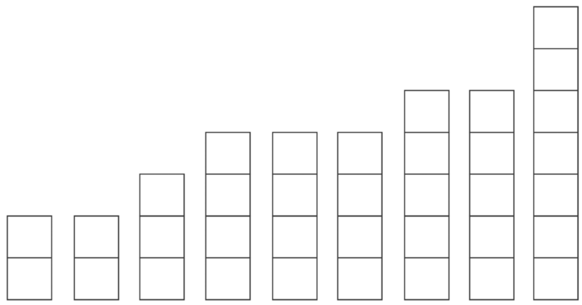


Figura 7. Policubos apilados y ordenados que representan el tamaño de las familias.

(1) Desconectar todos los policubos y redistribuirlos de uno en uno entre los nueve estudiantes hasta que se hayan repartido todos. En este caso, hay 36 policubos. Redistribuirlos entre 9 estudiantes da como resultado 9 pilas de 4 policubos cada una.

(2) Retirar un policubo de la pila más alta y colocarlo en una de las más bajas, y así sucesivamente hasta que todas las pilas estén niveladas.

Ambos métodos conducen a un mismo tamaño familiar de 4, que puede ser considerado un reparto equitativo o justo.

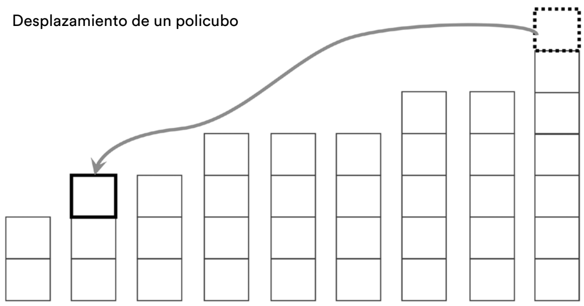


Figura 8. Desplazamiento de un policubo desde la pila más alta

Con el segundo enfoque, el alumnado empieza retirando un policubo de la pila más alta y colocándolo en una de las pilas más bajas. Así se consigue una nueva disposición de los cubos (véase la figura 8).

Este proceso se lleva a cabo hasta que todas las pilas están niveladas o casi niveladas — cuando hay un resto— (véase la figura 9).

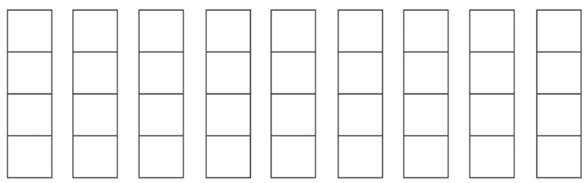


Figura 9. Policubos que representan el tamaño medio de una familia.

Tras el último movimiento, las 9 pilas están niveladas con 4 policubos cada una. Así se comprueba que 4 es un reparto equitativo para el tamaño de la familia. Es decir, si las nueve familias fueran de igual tamaño, el número de personas en cada hogar sería cuatro. Este valor del reparto equitativo es la media de la distribución. La media es un resumen numérico de la distribución de una variable cuantitativa usada para medir el centro de la distribución. La distribución de una variable cuantitativa resume los diferentes valores de la variable y la frecuencia con la que se da cada uno. La mayoría del alumnado y del profesorado saben cómo calcular la media mediante la suma de todas las observaciones y la división por el número de observaciones. Pero ¿qué le dice la media al alumnado sobre la distribución? ¿Cómo se espera que interprete la media? ¿Cómo se espera que describa la variabilidad de una distribución en relación con su media?

Supongamos que otros dos grupos de nueve estudiantes de la clase obtuvieron un valor de reparto equitativo de 6. ¿Qué otras representaciones podrían haber llevado a cabo con policubos?

Para responder esto, el alumnado debe darse cuenta, primero, de que tiene que empezar con 54 policubos. A continuación, puede crear dos distribuciones diferentes del tamaño de las familias, en las que el valor del reparto equitativo sea 6. Por ejemplo, consideremos los dos grupos siguientes: el grupo 1 (véase la figura 10) y el grupo 2 (véase la figura 11), con datos de 9 tamaños familiares de la clase, donde la cantidad de miembros para cada grupo es 6.

Dado que el valor del reparto equitativo para cada grupo es 6, no es posible establecer una diferencia entre los dos grupos a partir de este valor. Una pregunta de análisis podría ser:

¿Qué grupo está más cerca de tener un reparto equitativo?

El alumnado puede dar diferentes respuestas a esta pregunta, entre ellas:

(1) El grupo 2, porque este grupo tiene la frecuencia más alta de pilas de 6 policubos.
(2) El grupo 1, porque en este grupo haría falta mover menos policubos para nivelar todas las pilas al valor del reparto equitativo de 6.

La segunda respuesta, la de tener menos policubos que mover, se puede entender como contar el «número de pasos para igualar» o, en otras palabras, cuántos pasos en total harían falta para mover los policubos hasta crear grupos de igual tamaño. Un número menor de pasos indica que la distribución está más cerca de ser uniforme y que tiene menos variabilidad respecto a la media. El alumnado puede seguir el proceso para ver que, en el grupo 1, se tendrían que mover 8 cubos en un total de 48 pasos. En el grupo 2, haría falta mover 9 cubos en un total de 58 pasos. Por tanto, el grupo 1 está más cerca de estar igualado y tiene menos variabilidad respecto a la media que el grupo 2.

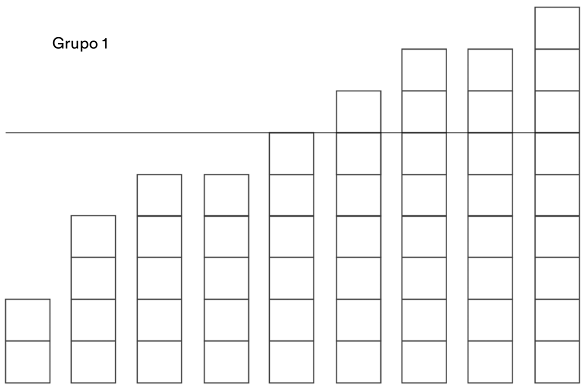


Figura 10. Disposición del grupo 1 con promedio de 6

Interpretar los datos

A continuación, el alumnado debe interpretar los resultados para responder la pregunta de investigación estadística original:

¿Cuál es el tamaño de la familia típica del alumnado de la profesora López?

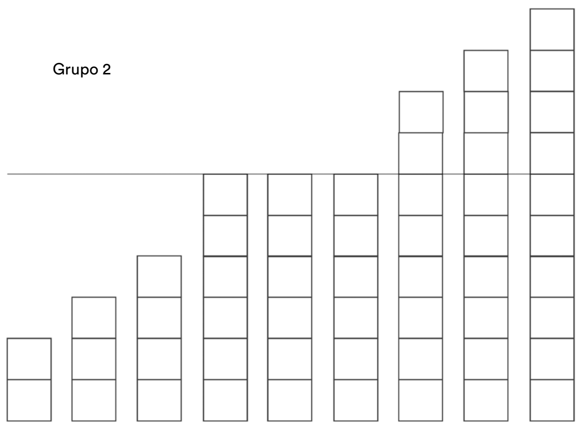


Figura 11. Disposición del grupo 2 con promedio de 6

A partir de los resultados de los dos últimos grupos, el alumnado puede comentar que, si todas las familias fueran del mismo tamaño, el número de personas en cada vivienda sería 6, que sería el reparto equitativo o valor medio. El grupo 1 tiene tamaños familiares más próximos a la media de 6, según el número de pasos necesarios para nivelar la distribución. El grupo 1 tiene menos variabilidad que el grupo 2. Esto ilustra que, al observar la distribución del tamaño de las familias, es importante no confiar en un único número de resumen (p. ej., la media). También debemos conocer cuánto varía el tamaño de las familias con respecto a la media para describir la representatividad de esos tamaños.

En el nivel A se investiga y se comprende cómo interpretar la media como el valor de reparto equitativo/justo y cómo cuantificar la variabilidad con respecto a la media como el número de pasos hasta el valor del reparto equitativo. En el nivel B se avanza hacia la interpretación de la media como el punto de equilibrio, y la cuantificación de la variabilidad con respecto a la media como la desviación absoluta media[11] (DAM).

Ejemplo 3: Qué aspecto tienen las mariquitas. Recogida, resumen y comparación de datos

Formular preguntas de investigación estadística

Un tema típico en Ciencias del nivel A es estudiar la estructura y la función de las partes del cuerpo de un organismo, así como su entorno. Actualmente, los estándares curriculares en Ciencias incluyen numerosos conceptos estadísticos y abren un espacio en el currículo Pre-K–12 en el que introducir el pensamiento estadístico. Por ejemplo, consideremos una lección de Ciencias en la que se estudie qué hacen las mariquitas y qué aspecto tienen (véase la figura 12). Con la orientación y la ejemplificación del profesorado, la clase formula tres preguntas de investigación estadística:

¿Qué aspecto suelen tener las mariquitas?

¿Cuántas manchas tienen generalmente las mariquitas?

¿Suelen tener más manchas las mariquitas rojas que las negras?

Recoger/considerar los datos

Los datos se proporcionan en forma de imágenes de mariquitas. Al ver las imágenes, el alumnado puede darse cuenta inmediatamente de que hay variación en el número de manchas de las mariquitas y en su color, y puede registrar información sobre el número de manchas y el color de cada mariquita de las imágenes, o sobre otras características que considere que pueden ser relevantes. El profesorado orienta al alumnado en la elaboración de preguntas de recogida de datos que deberán responderse sobre cada mariquita, a fin de empezar a contestar las preguntas de investigación estadística.

¿Cuántas manchas tiene la mariquita?

¿De qué color es la mariquita?

¿De qué color son las manchas de la mariquita?

El número de manchas de la mariquita es un ejemplo de variable numérica. Podemos obtener datos de variables numéricas mediante la toma de medidas (p. ej., estaturas o temperaturas del alumnado) o contando objetos (p. ej., el número de letras que tienen los nombres de los estudiantes, el número de bolsillos que hay en la ropa que llevan o el número de hermanos y hermanas de cada estudiante). Las variables numéricas también reciben el nombre de *variables cuantitativas*, término con el que se las denomina a lo largo de este documento.

Figura 12. *Hippodamia convergenes* (mariquita)

El color de la mariquita es un ejemplo de variable cualitativa. Los datos de las variables cualitativas se observan de acuerdo con su categoría; todas las categorías son mutuamente excluyentes y conjuntamente exhaustivas, lo que significa que no se superponen y que representan todas las observaciones posibles.

[11] N. de T.: La literatura también se refiere a la DAM como *desviación media absoluta*, *media de las desviaciones absolutas* o, simplemente, *desviación absoluta*.

Con la orientación del profesorado, el alumnado selecciona las categorías que se usarán en relación con el color: negro, naranja y rojo. Para cada una de las imágenes (en la figura 13 se muestra un ejemplo para 19 estudiantes), el alumnado planteará las preguntas de recogida de datos y registrará la información para las tres variables: 1) color del cuerpo; 2) número de manchas; y 3) color de las manchas (si es aplicable).

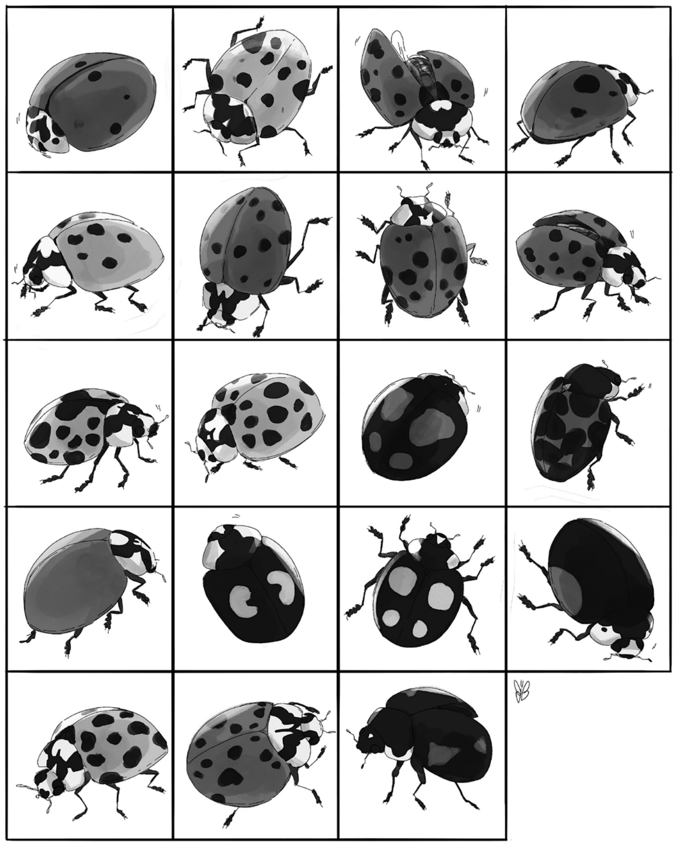


Figura 13. Ejemplos de las tarjetas de fotografías 4 × 4 (https://en.wikipedia.org/wiki/Harmonia_axyridis)

El alumnado debería darse cuenta de que las mariquitas son simétricas, por lo que, si cuenta el número de manchas que hay en uno de los lados, sabrá cuántas hay en el otro. Se registra el número total de manchas. A veces, los datos están desorganizados o no son claros. Por ejemplo, en este caso, algunas manchas están muy desdibujadas y no parecen una mancha. Es importante que la clase decida qué contará como mancha (p. ej., si se contarán todas las marcas que tengan alguna forma y todas las manchas que se encuentren en el margen del élitro); de este modo, se reducirán los errores de medida introducidos por el alumnado al registrar la información sobre las mariquitas observadas. En el nivel A se debe comprender la importancia de recoger los datos de manera consistente.

En este nivel, los datos recogidos se pueden registrar de diversas formas. Inicialmente, el alumnado podría considerar una variable y registrar los valores para cada mariquita, como en la tabla 4.

Tabla 4. Tabla de datos para el nivel A inicial

N.º mariquita	Color del cuerpo
1	R
2	A
...	...

En una etapa posterior del nivel A, puede registrar las tres variables a la vez; por ejemplo, registrando las respuestas a las tres preguntas para cada mariquita, como en la tabla 5, en la que se muestra una posible estructura tabular con la siguiente información en cada celda, de izquierda a derecha: número de manchas, color del cuerpo (R = rojo, N = negro, A = naranja) y color de las manchas (R = rojo, N = negro, A = naranja). Cada una de las celdas puede considerarse como una tarjeta de datos, una herramienta organizativa para los datos. En el ejemplo 6 se ofrecen detalles sobre las tarjetas de datos.

Tabla 5. Tabla de ejemplo de las tarjetas de datos de las mariquitas

6RN	10AN	16RN	6RN
10AN	16RN	18RN	18RN
20AN	20AN	4NR	16RN
0R-	2NR	4NR	2NR
20AN	16RN	4NR	

Al compartir sus estrategias como clase y comentarlas, el alumnado puede comenzar a reconocer la importancia de tener una estrategia que le permita organizar los datos de una manera útil.

Con el tiempo, el alumnado debería tratar de crear un método más productivo de organizar los datos. Quienes tengan más experiencia deben ser capaces de crear una tabla de datos en la que cada observación es una fila aparte (véase la tabla 6 como ejemplo). Esto podría hacerse utilizando papel y lápiz, o mediante herramientas tecnológicas.

Tabla 6. Tabla de datos de las tarjetas de mariquitas

N.º mariquita	Número de manchas	Color del cuerpo	Color de las manchas
1	6	R	N
2	10	A	N
...	...	...	...

Analizar los datos

Al inicio del nivel A, el alumnado puede utilizar un gráfico de recuento icónico para analizar los datos, porque le permite llevar la cuenta de qué mariquita está siendo representada. A medida que se avanza en el nivel A, el gráfico de puntos[12] es una representación gráfica apropiada para una variable cuantitativa. Con la orientación del profesorado, el alumnado debe ser capaz de asociar una mariquita con un punto del gráfico. Esta conexión es importante, pues el gráfico de puntos ya no permite distinguir cada mariquita individualmente. Los gráficos de puntos se pueden hacer a mano o mediante recursos tecnológicos; el eje horizontal representa, normalmente, las categorías de la variable estudiada. Para comparar el número de manchas que tienen las mariquitas de diferentes colores, el alumnado del nivel A puede utilizar varios gráficos de puntos con la misma escala y apilados, uno encima del otro (véase la figura 14 como ejemplo).

Utilizando un único gráfico de puntos o varios desglosados por los diferentes colores de las mariquitas, el alumnado puede responder una serie de preguntas de análisis sobre la variable cuantitativa «número de manchas»; por ejemplo:

¿Qué número de manchas es el más común/típico en todas las mariquitas?
¿Y solo en las rojas?
¿Y en las naranjas?
¿Y en las negras?

¿Cuál es el número de manchas más bajo/más alto en las mariquitas?
¿Y solo en las rojas?
¿Y en las naranjas?
¿Y en las negras?

Se debe ayudar al alumnado del nivel A para que piense sobre la distribución de una variable cuantitativa y la variabilidad en sus valores. El alumnado debe entender que la mediana representa el valor que está en la mitad o centro de la distribución de una variable cuantitativa. El mismo número de datos (aproximadamente la mitad) es mayor que la mediana y menor que la mediana. Las medianas de la figura 14 son: 14 manchas para las mariquitas rojas, 18 para las naranjas y 4 para las negras.

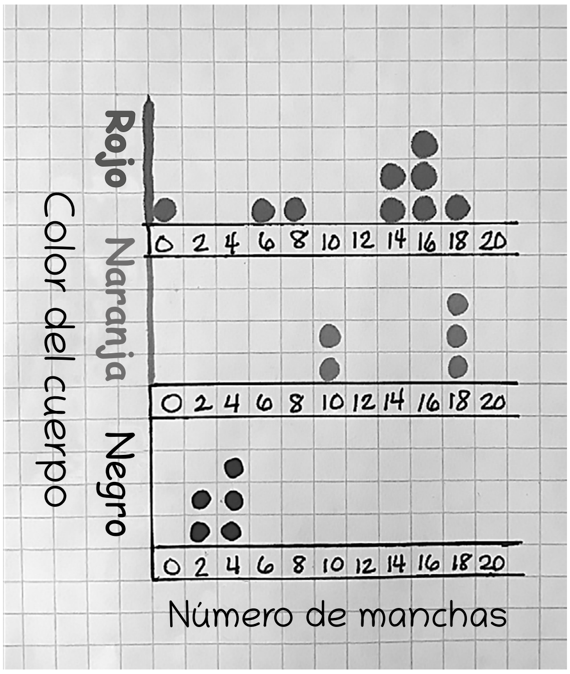


Figura 14. Gráficos de puntos hechos a mano y apilados para el recuento de manchas en mariquitas de distinto color

Para ilustrar la media como punto central, el alumnado puede crear un gráfico que represente el número de manchas que hay en la tarjeta de mariquitas que sujetan en la mano. Quienes tengan imágenes en las que haya dos manchas se colocan en una fila; quienes tengan imágenes con cuatro manchas se sitúan en una fila paralela contigua; a continuación, quienes tengan imágenes con seis manchas, y así sucesivamente hasta que toda la clase esté dispuesta en filas. Una vez se haya organizado a todo el grupo, el profesorado puede pedirle a una persona de cada uno de los dos extremos del gráfico que se siente, e ir repitiendo este proceso hasta

[12] N. de T.: En el original se usa *drawn stacked dotplot* cuando se trata de una representación estadística de los datos, dibujada manualmente, en la que se utilizan puntos como marca de recuento (véase la figura 14); y *stacked dotplot* cuando la representaciós no se dibuja manualmente (véase la figura 16).

que solo quede una persona en pie, que será quien represente la mediana. Hasta que el alumnado domine la idea de un punto medio, es mejor usar un número impar de datos para que la mediana sea clara.

0		1/2		1
Imposible	Improbable o menos probable	Igual de probable que ocurra y que no ocurra	Probable o más probable	Seguro

Figura 15. Escala de probabilidades

Las preguntas de análisis que requieren realizar aseveraciones sobre la probabilidad de ciertos enunciados permiten que, en el nivel A, se conecte lo que se está viendo en el caso concreto de una clase con un panorama más amplio. El alumnado puede aprender a asignar números, de manera informal, a la probabilidad de que algo ocurra. En la figura 15 se muestra un ejemplo de asignación de números en una recta numérica.

A partir de estos números se podrían plantear al alumnado las siguientes preguntas de análisis sobre los datos en cuestión:

¿Cómo de probable es encontrar una mariquita negra que tenga más de cinco manchas? ¿Y menos de cinco manchas?

¿De qué color son las mariquitas en las que varía más el número de manchas? ¿Y menos?

Si sacara una de las imágenes de mariquitas de un sombrero y tuviera seis manchas, ¿de qué color crees que sería?

¿Es más probable que las mariquitas naranjas tengan más de diez manchas o que tengan menos?

Si sacara una de las tarjetas de mariquitas de un sombrero y fuera una mariquita negra, ¿qué probabilidad hay de que tuviera dos manchas?

Interpretar los resultados

Utilizando sus análisis, el alumnado puede contestar la pregunta de investigación estadística —«¿Qué aspecto suelen tener las mariquitas?»— con una respuesta como la siguiente:

«A partir de nuestras imágenes, las mariquitas de la colección eran una mezcla de mariquitas rojas, naranjas y negras. Las mariquitas rojas tenían entre 0 y 18 manchas negras. Las mariquitas naranjas tenían entre 10 y 20 manchas negras. Las mariquitas negras eran diferentes, porque tenían o 2 o 4 manchas con una mezcla de colores».

Una estudiante podría escribir lo siguiente para responder la pregunta «¿Cuántas manchas tienen las mariquitas en nuestra colección de tarjetas?»:

«Las mariquitas rojas tienen entre 0 y 18 manchas. El número de manchas más común es 16. El promedio de manchas para las mariquitas rojas es 14. Las mariquitas naranjas tienen entre 0 y 18 manchas. El número de manchas más común es 18. El promedio de manchas para las mariquitas rojas es 18. Este valor es un poco mayor que el de las mariquitas rojas. Las mariquitas negras tienen 2 o 4 manchas. De las cinco mariquitas, dos tenían 2 manchas y el resto tenían 4».

Para responder la pregunta comparativa —«¿Suelen tener más manchas las mariquitas rojas que las negras?»—, el alumnado puede responder:

«Las mariquitas rojas tienen entre 0 y 18 manchas. El número de manchas más común es 16. El promedio de manchas para las mariquitas rojas es 14».

«Las mariquitas negras solo tienen 2 o 4 manchas. La mediana del número de manchas para las mariquitas negras es 4. Solo hay una mariquita roja que tenga menos de 4 manchas; hay una que no tiene ninguna mancha. Todas las otras mariquitas rojas tienen 6 o más manchas. Este análisis sugiere que las mariquitas negras de nuestras imágenes suelen tener menos manchas que las mariquitas rojas.»

Ejemplo 4: El cultivo de judías. Un experimento comparativo simple

El experimento simple, que consiste en tomar medidas bajo una determinada condición o un conjunto de condiciones, es otro tipo de diseño de recogida de datos apropiado para el nivel A. Un experimento comparativo simple es como un experimento científico en el que se comparan los resultados bajo dos o más condiciones. Por ejemplo, el alumnado puede sembrar alubias (semillas de las judías) en dos condiciones de cultivo diferentes y dejarlas germinar, y comparar después qué grupo de plantas crece más rápido durante un periodo de tiempo: las que están expuestas a la luz o las que están a oscuras. Los datos recogidos conducen a una investigación de experimento comparativo. También puede recoger los datos de crecimiento de cada una de las plantas a lo largo del mismo periodo de tiempo, lo que le permitiría hacer una investigación de series temporales.

Formular preguntas de investigación estadística

En la situación del experimento comparativo, una pregunta de investigación estadística podría ser:

Tras 15 días, ¿tienden a ser más altas las plantas cultivadas expuestas a la luz que las cultivadas a oscuras?

Con las preguntas comparativas de investigación estadística, la pregunta de investigación estadística planteada se formula como una hipótesis (o conjetura) para indicar el grupo que la persona a cargo de la investigación cree que será más alto. En este caso, la hipótesis es que las judías cultivadas expuestas a la luz crecerán más que las cultivadas a oscuras.

Recoger/considerar los datos

En este nivel A, el alumnado decidirá qué judías serán cultivadas en el ambiente iluminado y cuáles crecerán en el ambiente oscuro. El objetivo del experimento es la comparación entre estas dos condiciones. El tipo de entorno de iluminación es un ejemplo de variable cualitativa. Las medidas de la altura de las plantas, que representa una variable cuantitativa continua, pueden tomarse tras un tiempo prefijado. Estas medidas pueden responder la pregunta de investigación estadística de si hay unas condiciones de iluminación que sean mejores para el cultivo de las judías. Cuando se realizan experimentos, el profesorado debe establecer unos criterios con el alumnado acerca de cómo manejar ciertas situaciones que puedan surgir. Por ejemplo, algunas semillas podrían no germinar o algunas plantas podrían morir, y el alumnado debe decidir cómo registrar estas circunstancias en su recogida de datos. El concepto de experimento se desarrolla de forma más completa en el nivel C.

Analizar los datos

El alumnado del nivel A puede registrar la altura (en centímetros) de las judías cultivadas a oscuras y la de las cultivadas expuestas a la luz mediante un gráfico de puntos. Las alturas del día 8 están representadas en la figura 16.

Seguidamente, hace afirmaciones sobre lo que observa en el gráfico de puntos. Por ejemplo:

«He visto que la mediana de la altura de las judías cultivadas a oscuras es 2,5 cm, mientras que la mediana de la altura de las judías cultivadas expuestas a la luz es 7,5 cm».

A medida que el alumnado progresa en el nivel A, se le puede presentar la idea de la media como reparto equitativo (véase el ejemplo 2). La media como medida de centralización se desarrolla

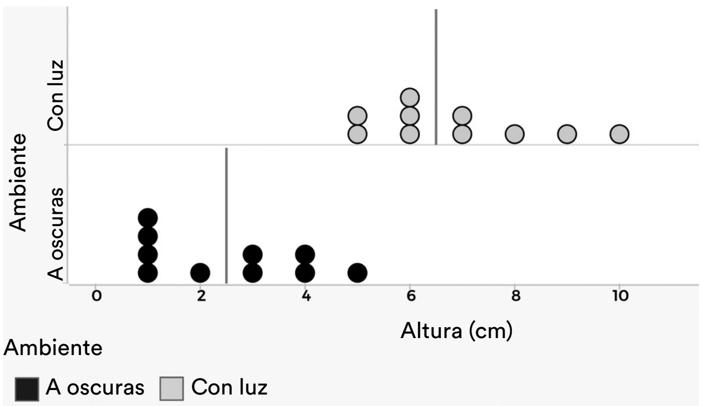


Figura 16. Gráficos de puntos apilados para las alturas de las judías en condiciones de luz y de oscuridad

en mayor detalle en los niveles B y C. La media y la mediana son medidas de centralización para describir el centro de una variable cuantitativa.

También es útil saber cómo varían los datos a lo largo del eje horizontal. El rango, que es la diferencia entre los valores máximo y mínimo, es una medida de la variabilidad para una distribución. El rango sólo tiene sentido con datos de una variable cuantitativa. La variabilidad de la distribución también puede describirse proporcionando los valores más alto y más bajo. Por ejemplo, la altura de las judías cultivadas a oscuras varía entre 1 cm y 5 cm, mientras que la de las judías cultivadas expuestas a la luz varía entre 5 cm y 10 cm. Así se aporta más información para describir una distribución que indicando únicamente los rangos (p. ej., decir que el rango de las alturas de las judías cultivadas a oscuras es de 4 cm y que el rango de las alturas de las judías cultivadas expuestas a la luz es de 5 cm), ya que se proporcionan medidas de posición, además de la variabilidad.

Buscar agrupaciones y espacios vacíos en la distribución ayuda al alumnado a identificar la forma de la distribución. En el nivel A se deben desarrollar algunas nociones sobre por qué una distribución adquiere una forma particular en el contexto de la variable considerada.

- → *¿Tiene la distribución una agrupación (o pico) que destaca más, con grupos más pequeños de tamaño similar a cada lado?* En caso afirmativo, la distribución puede describirse como *simétrica*.
- → *¿Tiene la distribución una agrupación más destacada, con grupos más pequeños a cada lado que no son de igual tamaño?* El alumnado puede clasificar esta distribución como «desequilibrada» o puede usar el término *asimétrica*.
- → *¿Por qué toma la distribución esta forma?* Utilizando el gráfico de puntos de la figura 16, el alumnado se dará cuenta de que tanto las judías cultivadas a oscuras como las cultivadas expuestas a la luz tienen distribuciones «desequilibradas», con la principal agrupación en el extremo izquierdo de las distribuciones y unos pocos valores a la derecha.

A medida que se progresa al nivel B, la consideración de la forma de una distribución conducirá a comprender qué medidas son apropiadas para describir el centro y la variabilidad.

Interpretar los resultados

El análisis revela que la altura media de las judías cultivadas expuestas a la luz excede en 5 cm la altura media de las otras judías. Casi todas las judías cultivadas expuestas a la luz son más altas que las cultivadas a oscuras. De hecho, las judías que se superponen son las más bajas de las cultivadas con luz y las más altas de las otras. Por lo tanto, las plantas de nuestro experimento en el ambiente iluminado tienden a ser más altas que las plantas en el ambiente a oscuras.

Ejemplo 5: El cultivo de judías (continuación). Series temporales

Formular preguntas de investigación estadística

En el nivel A, el alumnado puede analizar la altura de las distintas plantas de judías durante un periodo de 15 días. En esta situación, la pregunta de investigación estadística podría ser:

¿Cómo crece una planta de judías en un ambiente iluminado durante 15 días?

Recoger/considerar los datos

Para el experimento comparativo, el alumnado sólo tenía que registrar la altura al término de los 15 días; en cambio, en la recogida de datos para la serie temporal, debe registrar la altura cada día, con el objetivo de hacer un seguimiento del crecimiento.

Los datos de una planta recogidos por un estudiante pueden registrarse en una tabla (véase la tabla 7).

El alumnado puede ver que en la tabla falta el día 9, y es importante tratar esta cuestión antes de analizar los datos.

¿Por qué falta información y cómo puede afectar esto al análisis de los datos? Como no pudieron acceder a las plantas el día 9 para tomar medidas, decidieron que, puesto que tenían medidas de los seis días posteriores, los datos perdidos no afectarían a su capacidad de observar una posible tendencia.

Tabla 7. Planta de judía cultivada expuesta a la luz, crecimiento diario

Planta de judía A, crecimiento con exposición a la luz	
Número de día	**Altura (cm)**
1	0,75
2	1
3	1
4	2,2
5	4
6	6
7	6,5
8	7
9	N/A
10	10
11	10
12	12
13	14
14	14
15	16

Analizar los datos

En el nivel A, el alumnado puede usar un gráfico de series temporales, o gráfico temporal,[13] para analizar cómo cambia una variable cuantitativa a lo largo del tiempo.

En el gráfico temporal de la figura 17, el alumnado puede observar lo siguiente:

«La altura de la planta de judías aumentó cada día excepto del día 2 al 3 y del día 10 al 11.»

«El crecimiento de la planta de judías comenzó siendo bastante lento. Del día 1 al 3, la planta apenas creció, menos de un centímetro; y, transcurridos 15 días, creció hasta los 16 cm.»

«Lo máximo que creció la planta en un solo día fue 2 cm. Este crecimiento se produjo tres veces: entre los días 5 y 6, entre los días 11 y 12 y entre los días 14 y 15.»

El alumnado de este nivel debe ser consciente de cómo afectaría a la percepción del crecimiento un cambio en la escala del eje vertical, y darse cuenta de que aumentar los espacios entre las divisiones del eje de la altura podría hacer que pareciera que las judías crecieron más rápido de lo que realmente lo hicieron.

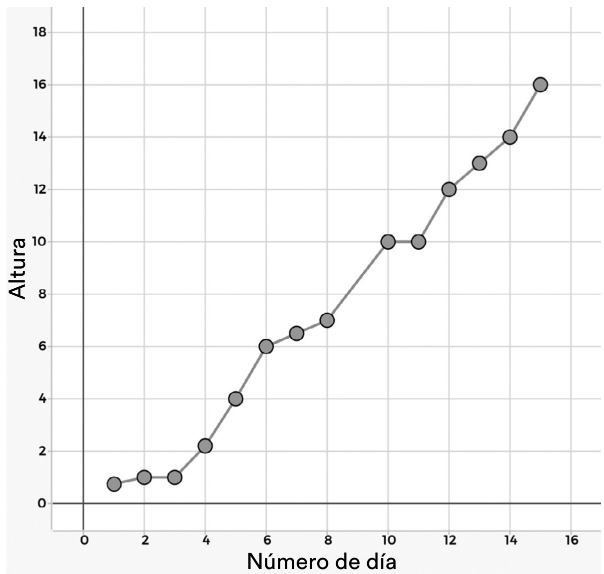


Figura 17. Gráfico temporal del crecimiento de una judía cultivada expuesta a la luz durante 15 días

Interpretar los resultados

El análisis muestra que la planta de judías creció de 0,75 cm a 16 cm en el periodo de 15 días. La tasa de crecimiento fue aceptablemente consistente a lo largo de los 15 días. El alumnado puede comparar sus resultados con otros obtenidos bajo las mismas condiciones de cultivo, para examinar las tasas de crecimiento típicas dentro de cada grupo y desarrollar futuras preguntas de investigación estadística.

[13] N. de T.: En el original se utiliza el término *timeplot*.

Ejemplo 6: Census at School. Uso de datos secundarios y búsqueda de asociaciones

Al plantear preguntas de investigación estadística, en lugar de recoger los datos personalmente, el alumnado puede encontrarse con un conjunto de datos secundarios para analizar e interpretar a fin de responder las preguntas. Los datos secundarios son datos que han sido recogidos con anterioridad. El proceso de investigación estadística también puede empezar directamente por la consideración de estos datos. Por ejemplo, un estudiante puede considerar un conjunto de datos secundarios y, después, plantear preguntas de investigación estadística sobre ese conjunto de datos, como se ilustra en este ejemplo.

Recoger/considerar los datos

Consideremos un conjunto de datos secundarios recogidos de la clase de 5.º del profesor Johnson, en una escuela de primaria, utilizando la encuesta *Census at School* (ASA, s.f.-a). Census at School es un proyecto internacional de aula para estudiantes desde 4.º de primaria hasta 2.º de bachillerato que se centra en la resolución de problemas estadísticos utilizando sus propios datos reales.

La encuesta planteaba una serie de trece preguntas, de las que se han escogido cinco para analizarlas aquí. Las respuestas de cada estudiante se presentan en una tarjeta de datos; es decir, una tarjeta o papel que muestra, para cada individuo, los valores de las variables incluidas en el conjunto de datos. Las tarjetas de datos son herramientas que permiten visualizar múltiples medidas registradas en la misma unidad de observación. Estas herramientas ayudan al alumnado a reconocer que hay muchos atributos que se pueden medir en una misma unidad de observación, como puede ser un estudiante de la clase de 5.º curso. En la figura 18 se presenta un ejemplo de la tarjeta de datos con las respuestas de una persona (izqda.), así como la leyenda de las tarjetas de datos (dcha.).

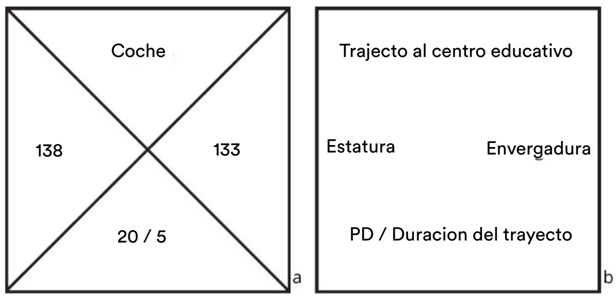


Figura 18. Ejemplo de una tarjeta de datos de una persona (izqda.) y leyenda que muestra las posiciones de las cinco variables (dcha.)

La tarjeta de datos incluye información sobre la forma en que cada estudiante suele ir al centro educativo (parte superior de la tarjeta); su estatura en centímetros (parte izquierda de la tarjeta), su envergadura en centímetros (parte derecha de la tarjeta); la longitud, en centímetros, de su pie derecho (valor inferior izquierdo de la tarjeta); y el tiempo, en minutos, que le lleva ir al centro educativo (valor inferior derecho de la tarjeta).

El alumnado debe examinar la tarjeta de datos y ver cuál es la unidad de medida para cada variable. Por ejemplo, el alumnado de los Estados Unidos se daría cuenta de que las unidades no son las habituales en su país, sino las del sistema métrico decimal.[14] Estas tarjetas de datos incluyen cuatro variables cuantitativas y una cualitativa.

Esta tarjeta de datos en concreto es de una estudiante que va a la escuela en coche, mide 138 cm, tiene una envergadura de 133 cm y una longitud de pie derecho (PD) de 20 cm, y tarda 5 minutos en llegar al centro educativo (duración del trayecto). Las tarjetas de datos que se muestran en la figura 19 son un subconjunto del conjunto total de datos y pueden encontrarse en *More 4 U* (NCTM, s.f.) junto con las instrucciones de preparación.

Cuando el alumnado del nivel A dispone del conjunto de tarjetas de datos de su clase, puede examinar en detalle los datos para identificar y comprender el contexto de cada variable. A continuación, puede plantear preguntas de investigación estadística interesantes. *More 4 U* contiene un ejemplo de cómo presentar las tarjetas de datos para fomentar que el alumnado examinen los datos en detalle a fin de identificar sus cinco variables.

[14] N. de T.: Al contrario que en la mayoría de los países, en los Estados Unidos no es común el uso del sistema métrico decimal. El alumnado de países donde es habitual el sistema métrico decimal no apreciaría nada extraño en estas unidades de medida.

Formular preguntas de investigación estadística

Hay varias preguntas de investigación estadística que pueden responderse utilizando estos datos, entre ellas, las siguientes:

¿Cuál es la estatura del alumnado de esta clase de 5.º curso? ¿Cuánto tardan en llegar al centro educativo quienes van en bus? (preguntas de investigación estadística de resumen)

¿Suelen tardar más en llegar al centro educativo los miembros de esta clase que van en bus que los que van a pie? (pregunta de investigación estadística de comparación)

¿Existe alguna relación entre la estatura y la envergadura del alumnado de esta clase de quinto curso? (pregunta de investigación estadística de asociación)

Esta investigación estadística se centrará en la pregunta de investigación estadística de asociación entre la envergadura y la estatura.

Trayecto al centro educativo; Estatura, Envergadura; PD / Duración del trayecto	Coche; 138, 133; 20 / 5	A pie; 162, 133; 21 / 3	A pie; 140, 134; 22 / 15	Coche; 150, 156; 23 / 12
A pie; 141, 132; 21 / 6	Coche; 131, 131; 24 / 10	Bus; 138, 140; 22 / 10	Coche; 127, 123; 20 / 5	Coche; 135, 140; 21 / 7
Bus; 146, 156; 23 / 5	Coche; 138, 135; 21 / 60	A pie; 143, 135; 23 / 4	Otro; 145, 157; 25 / 4	Coche; 140, 157; 21 / 5
Coche; 142, 130; 21 / 20	Coche; 146, 143; 23 / 5	Coche; 156, 157; 23 / 1	Coche; 151, 147; 24 / 15	Coche; 148, 162; 24 / 9
A pie; 144, 144; 23 / 1	Coche; 148, 138; 20 / 3	Bus; 143, 138; 22 / 15	Coche; 142, 127; 22 / 11	Bus; 140, 136; 19 / 5
A pie; 143, 133; 22 / 7	Bus; 148, 149; 23 / 35	A pie; 148, 153; 22 / 10	Bus; 145, 151; 23 / 7	Bus; 153, 142; 24 / 25
Coche; 153, 149; 24 / 9	Coche; 158, 163; 25 / 17	A pie; 145, 163; 26 / 20	Coche; 151, 152; 24 / 5	Coche; 148, 139; 19 / 10

Figura 19. Treinta y cuatro tarjetas de datos de la clase de 5.º de primaria del profesor Johnson

Analizar los datos

Cuando los valores de dos variables cuantitativas provienen del mismo individuo u objeto, el diagrama de dispersión sirve para representar gráficamente los datos. En el nivel A, el foco de atención debe ser la interpretación del diagrama de dispersión, no su creación. El alumnado puede consultar un diagrama de dispersión ya elaborado, con las estaturas y las envergaduras correspondientes al conjunto de datos proporcionado (véase la figura 20), con el objetivo de buscar una relación entre estas dos variables cuantitativas.

En el nivel A, el alumnado puede usar el diagrama de dispersión para hacer una búsqueda informal de tendencias y patrones. Asimismo, puede preguntarse si existe una asociación entre la estatura y la envergadura; y, en caso afirmativo, advertir cómo de fuerte podría ser. Si los gráficos son creados por *software*, también puede fijarse en si los ejes empiezan en cero y cuáles son los incrementos.

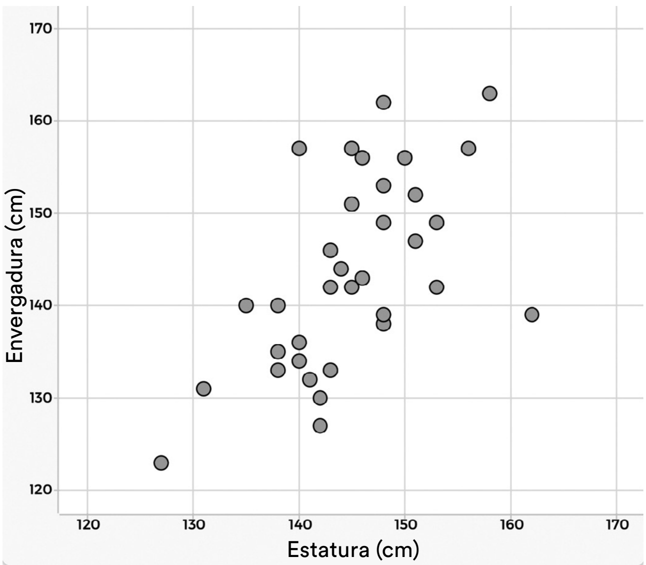


Figura 20. Diagrama de dispersión de la estatura y la envergadura

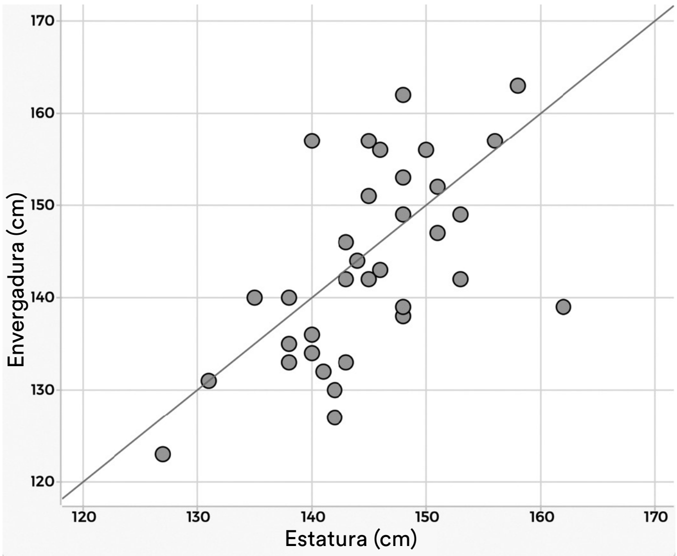


Figura 21. Diagrama de dispersión de la estatura y la envergadura con una posible recta a ojo

El alumnado debe ser capaz de describir la relación entre las dos variables. En el diagrama de dispersión de la figura 21 —donde se muestra la envergadura frente a la estatura—, generalmente, a medida que una variable crece, también lo hace la otra. Hacia el final del nivel A puede examinar si esta afirmación es cierta para su clase (p. ej., si la estatura de las personas tiende a estar relacionada con su envergadura). Para ello, puede estimar una recta que se ajuste a los datos observándolos en el diagrama de dispersión y, después, trazando la recta.

A partir de estos datos, el alumnado del nivel A puede ver que hay estudiantes cuya estatura es igual que su envergadura; son las personas que se encuentran sobre la recta, y 7 de los 34 estudiantes están sobre la recta o cerca de ella. Hay 10 personas cuya envergadura es mayor que su estatura; son quienes están por encima de la recta. Y 17 estudiantes tienen mayor estatura que envergadura. En su mayoría, la estatura y la envergadura de los estudiantes siguen el patrón de la recta, lo que sugiere que quienes tienen más estatura también tienen más envergadura. Una persona tiene una envergadura muy reducida (aproximadamente 139 cm) en comparación con su estatura (aproximadamente 162 cm). En el nivel A, el alumnado puede preguntarse si las medidas de esa persona se tomaron correctamente.

En el nivel B, estas tendencias y patrones se cuantificarán con medidas de correlación.

Interpretar los resultados

El análisis revela que la mayor parte del alumnado de la clase tiene estaturas similares a sus envergaduras. Las personas más altas tienden a tener envergaduras mayores. El alumnado del nivel A está expuesto a situaciones que requieren el desarrollo del lenguaje técnico para expresar lingüísticamente relaciones y patrones de datos, por lo que puede desarrollar el vocabulario técnico necesario para comunicar sus descubrimientos mediante la ejemplificación del uso apropiado de términos de resumen como *tiende a*, *típico*, *normalmente* y *similar*.

More 4 U incluye un ejemplo en que el alumnado del nivel A puede usar tarjetas de datos en las que hay ítems no tradicionales, lo que lo acerca tempranamente a la ciencia de datos.

Resumen del nivel A

Si el alumnado se familiariza con las ideas y los conceptos del nivel A, estará preparado para continuar desarrollando y mejorando su comprensión de los conceptos clave para la alfabetización estadística en el nivel B.

A través de las investigaciones estadísticas del nivel A se comienza a entender el papel de las preguntas en el proceso de resolución de problemas estadísticos, cómo pensar y representar múltiples variables a la vez, cómo representar gráficamente los datos de diferentes maneras y cómo pensar informalmente sobre la probabilidad. A medida que se avanza del nivel A al nivel B y al nivel C, es importante tener siempre en cuenta que entender y manejar la variabilidad es la esencia para desarrollar el sentido de los datos.

Nivel B

Introducción

Aspectos esenciales de cada componente

Ejemplo 1: Revisión del nivel A. La elección de la música para la fiesta escolar. Grupos mayores y multivariantes

Ejemplo 2: La elección de la música para la fiesta escolar (continuación). Comparación de grupos

Ejemplo 3: La elección de la música para la fiesta escolar (continuación). Conexión entre dos variables cualitativas

Ejemplo 4: Los pinzones de Darwin. Comparación de una variable cuantitativa entre grupos

Ejemplo 5: Los pinzones de Darwin (continuación). Separación frente a solapamiento

Ejemplo 6: Los pinzones de Darwin (continuación). La medida de la intensidad de la asociación entre dos variables cuantitativas

Ejemplo 7: Los pinzones de Darwin (continuación). Series temporales

Ejemplo 8: Dollar Street. Imágenes como datos

Ejemplo 9: Memoria y música. Experimentos comparativos

Resumen del nivel B

Introducción

La enseñanza en el nivel B debe construirse sobre los fundamentos estadísticos desarrollados en el nivel A y sentar las bases para la alfabetización estadística en el nivel C. Las actividades didácticas del nivel B siguen enfatizando el proceso de resolución de problemas estadísticos y reflejan el espíritu de una práctica estadística auténtica. En este nivel se adquiere más consciencia sobre cómo usar el cuestionamiento para guiar el razonamiento estadístico. También se presta mayor atención a los tipos de variables en los conjuntos de datos y a cómo esas variables pueden ayudar a abordar las preguntas de investigación estadística planteadas.

El alumnado sigue usando sus habilidades con los resúmenes gráficos, tabulares y numéricos introducidos en el nivel A, pero las amplía para investigar problemas más sofisticados, como las posibles asociaciones entre variables. Asimismo, pasa de utilizar datos de poblaciones completas a trabajar con datos extraídos de muestras de poblaciones, realizar comparaciones entre dos grupos y observar los cambios a lo largo del tiempo. En el nivel B es posible que las muestras no sean aleatorias, por lo que es importante comprender las limitaciones y el alcance de los estudios. Por otra parte, se deja de trabajar exclusivamente con conjuntos pequeños de datos, por lo que el alumnado debe empezar a emplear la tecnología como ayuda para gestionar el proceso de resolución de problemas estadísticos.

Aspectos esenciales de cada componente

I. Formular preguntas de investigación estadística

- → Reconocer que las preguntas de investigación estadística pueden usarse para definir temas de investigación, y que se pueden plantear muchas preguntas de investigación estadística sobre cualquier tema.
- → Entender que las preguntas de investigación estadística tienen en cuenta el contexto y la variabilidad presente en los datos.
- → Plantear preguntas de investigación estadística de resumen, comparativas y de asociación sobre una población más amplia utilizando muestras extraídas de esa población.
- → Plantear preguntas de investigación estadística que requieran observar una variable a lo largo del tiempo.
- → Entender que existen diferentes tipos de preguntas en estadística: las que se usan para enmarcar una investigación, las que se usan para recoger datos y las que se usan para guiar el análisis y la interpretación.
- → Plantear preguntas de investigación estadística a partir de datos recogidos de fuentes en línea y páginas web, *smartphones*, dispositivos de actividad física, sensores y otros dispositivos actuales.

II. Recoger/considerar los datos

- → Entender que los datos son información recogida y registrada con un propósito y que se pueden organizar y almacenar en distintas estructuras (p. ej., hojas de cálculo).
- → Comprender que una muestra puede usarse para responder preguntas de investigación estadística sobre una población. Reconocer las limitaciones y el alcance de los datos recogidos mediante la descripción del grupo, o población, del que se han obtenido.
- → Entender que los datos pueden usarse para hacer comparaciones entre diferentes grupos en un único momento temporal o dentro del mismo grupo a lo largo del tiempo.
- → Reconocer que los datos pueden ser recogidos utilizando encuestas y mediciones, y desarrollar una actitud crítica al analizar los métodos de recogida de datos.
- → Entender que las variables cuantitativas pueden ser discretas o continuas.

- → Entender cómo examinar los datos en detalle para determinar de qué manera fueron recogidos, de quién se obtuvieron, qué tipos de variables contienen, cómo se midieron esas variables (incluyendo las unidades usadas) y sus posibles resultados.
- → Comprender que se pueden recoger datos (datos primarios) u obtener datos ya existentes de otras fuentes (datos secundarios).
- → Entender cómo se usa la asignación aleatoria en experimentos comparativos para controlar las características que podrían afectar a las respuestas.

III. Analizar los datos

- → Representar la variabilidad de variables cuantitativas utilizando las representaciones adecuadas (p. ej., gráficos de puntos, gráficos de caja).
- → Aprender a utilizar las características clave de las distribuciones de variables cuantitativas, como:
  - o centrales (la media como el punto de equilibrio, y la mediana como el valor central de los datos ordenados);
  - o variabilidad (el rango intercuartílico y la desviación absoluta media [DAM]);
  - o forma (simétrica o asimétrica, y número de modas).
- → Usar el razonamiento sobre las distribuciones para comparar dos grupos a partir de variables cuantitativas.
- → Explorar patrones de asociación entre dos variables cuantitativas o cualitativas:
  - o medidas de correlación (razón de recuento de cuadrantes [RRC]);
  - o comparación de proporciones condicionadas entre variables cualitativas.

IV. Interpretar los resultados

- → Utilizar la evidencia estadística obtenida en los análisis para responder las preguntas de investigación estadística y comunicar los resultados mediante respuestas completas, con cierta orientación del profesorado.
- → Reconocer que los datos se pueden leer de una manera más profunda.
- → Generalizar a partir de la muestra proporcionando evidencia estadística para la generalización y mencionando la incertidumbre y la verosimilitud cuando sea necesario.
- → Reconocer la incertidumbre causada por la variabilidad entre las muestras.
- → Exponer las limitaciones de la información muestral (p. ej., una muestra puede ser o no representativa de una población más grande, variabilidad de la medición).
- → Comparar los resultados de un experimento obtenidos bajo condiciones diferentes.

Ejemplo 1: Revisión del nivel A. La elección de la música para la fiesta escolar. Grupos mayores y multivariantes

Formular preguntas de investigación estadística

En el nivel A, el alumnado quería seleccionar un grupo musical para la fiesta escolar de su curso mediante la realización de un censo en su clase en el que se respondiera la pregunta de investigación estadística:

¿Qué tipo de música le gusta al alumnado de nuestro curso?

Inicialmente, el alumnado recogió los datos de su clase para contestar esta pregunta. Cabe recordar que se consideraba que la clase era toda la población y que los datos se recogían para todos los miembros de esa población. A continuación, el alumnado se planteó la posibilidad de extender los resultados de su clase a todo el curso.

En el nivel B, una investigación similar podría incluir la planificación para todo el instituto. Esto implica reconocer que una sola clase puede no ser representativa de las preferencias del alumnado de un curso o de todo el instituto. En este nivel se pueden comparar las preferencias de una clase con las de otras clases del centro y explorar la siguiente pregunta de investigación estadística:

¿Qué tipo de música le gusta al alumnado de nuestro centro?

Recoger/considerar los datos

En el nivel B, el alumnado reflexiona sobre cómo recogerá y registrará los datos, así como de quién puede obtenerlos.

La pregunta de investigación estadística —«¿Qué tipo de música le gusta al alumnado de nuestro centro?»— se refiere a las preferencias del alumnado de todo el centro. En este caso, un plan de recogida de datos podría ser usar una sola clase (p. ej., la clase de Matemáticas de 1.º de secundaria) como muestra con el fin de tomar decisiones para todo el instituto; con esta opción, el alumnado debería comentar las limitaciones de la muestra que ha elegido. Otra opción sería que eligiera al azar a estudiantes de cada clase, o bien que seleccionara dos o tres clases y que el alumnado de esas clases respondiera una encuesta.

En el nivel B, el alumnado, al entender los posibles errores que hay que evitar en el diseño de la encuesta (como las redacciones ambiguas y las preguntas dirigidas), debe mejorar las preguntas de encuesta utilizadas en el nivel A y ofrecer más opciones de respuesta. Además, tiene que recoger datos sobre diversos aspectos de un tema, lo que anticipará las respuestas a otras preguntas de investigación estadística.

Por ejemplo, el alumnado puede plantear una serie de preguntas de encuesta que le permita profundizar sobre los tipos de música que le gustan. Tras recoger los datos, puede analizar si parece probable que exista una asociación entre los diferentes tipos de música que le gustan. Esta información podría influir en la elección de la música para la fiesta escolar.

P1. Marca «Sí» para todos los tipos de música que te gusten entre los siguientes. Marca «No» para cualquiera que no te guste.

	Sí	**No**
Rap	☐	☐
Rock	☐	☐
Country	☐	☐
R&B	☐	☐
Pop	☐	☐
Clásica	☐	☐
Alternativa	☐	☐

P2. ¿Cuál es tu tipo de música favorito?

☐ Rap
☐ *Rock*
☐ *Country*
☐ *R&B*
☐ Pop
☐ Clásica
☐ Alternativa
☐ Otro

P3. ¿Cuál es tu segundo tipo de música favorito?

☐ Rap
☐ *Rock*
☐ *Country*
☐ *R&B*
☐ Pop
☐ Clásica
☐ Alternativa
☐ Otro

P4. ¿Preferirías tener un grupo musical en vivo o un *DJ* en la fiesta de fin de curso?

☐ Grupo en vivo
☐ *DJ*

A medida que el alumnado avanza en el nivel B, puede diseñar su método de recogida de datos mediante el uso de tecnología. Las herramientas de encuestas en línea ofrecen formas de recoger datos para después descargarlos en una hoja de cálculo. Disponer de los datos en una hoja de cálculo permite empezar a analizarlos utilizando la tecnología. El alumnado también puede darse cuenta de que, cuando cumplimenta encuestas en línea, sus datos se recogen en un espacio centralizado para que otras personas los analicen e interpreten.

Analizar los datos

Muchos de los resúmenes gráficos, tabulares y numéricos presentados en el nivel A se pueden mejorar y utilizar para análisis más sofisticados en el nivel B. En esta segunda etapa es frecuente que se representen, a la vez, múltiples variables o se empleen múltiples formas de representación para responder la pregunta de investigación estadística. Para analizar los datos de la encuesta recogidos usando una clase como muestra del centro, el alumnado del nivel B puede representar gráficamente el número de estudiantes a quienes les gusta cada tipo de música (véase la figura 22). Este gráfico de barras utiliza las respuestas de los estudiantes a la pregunta de encuesta 1 indicada anteriormente, en la que cada tipo de música es una categoría de la variable.

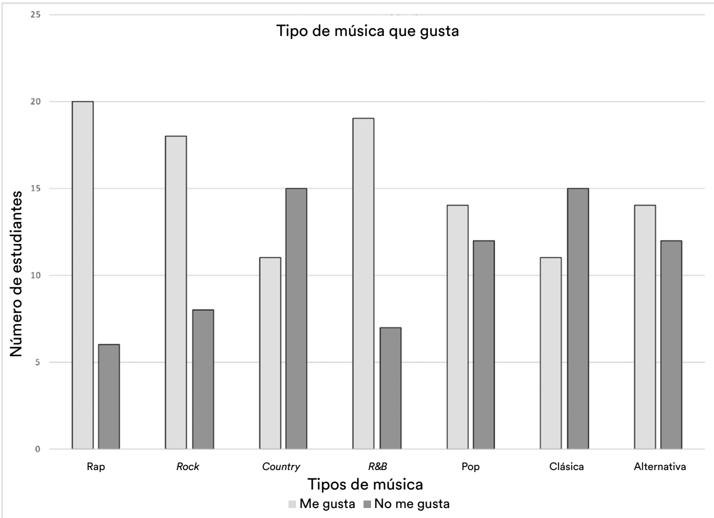


Figura 22. Gráfico de barras en paralelo para gustos musicales

El gráfico de barras muestra la frecuencia del número de estudiantes a quienes les gusta o no les gusta cada tipo de música. A partir de este gráfico, el alumnado puede ver que el rap es el género musical para el que más estudiantes han respondido «Sí», lo que se muestra en el gráfico como «Me gusta» (20 de 26 estudiantes). El *R&B* es el segundo que más gusta, seguido del *rock*. También hay más estudiantes a quienes les gustan el pop y la música alternativa que estudiantes a quienes no les gustan. Respecto al *country* y la música clásica, hay más estudiantes que dicen «No», lo que se muestra en el gráfico como «No me gusta», que los que dicen «Sí». El gráfico sugiere que el rap, el *R&B* y el *rock* son los tipos de música más populares entre la clase.

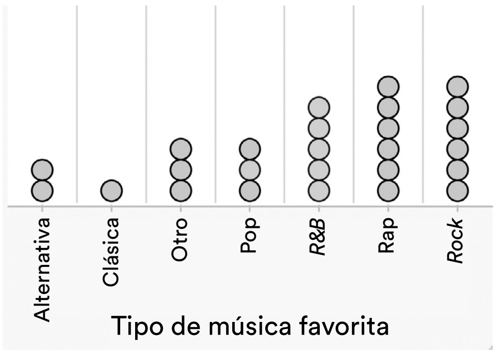


Figura 23. Música favorita según tipo

Se pueden analizar las respuestas a la pregunta de encuesta 2 para ver cuál es el tipo de música favorito del alumnado. La figura 23 muestra que hay el mismo número de estudiantes (6) que identificaron el rap y el *rock* como su música favorita.

El alumnado puede analizar la música favorita de la clase y la segunda más preferida (respuestas a las preguntas de la encuesta 2 y 3) mediante un gráfico de doble entrada (véase la figura 24). Aquí puede ver que toda la clase, excepto dos personas, ha escogido el *rock*, el rap o el *R&B* como su primer y/o segundo tipo de música favorito. En este gráfico, cada punto representa a una persona de la clase que ha contestado la encuesta. Este gráfico de doble entrada muestra 49 casillas (7 posibles tipos de música como primera opción y otros 7 como

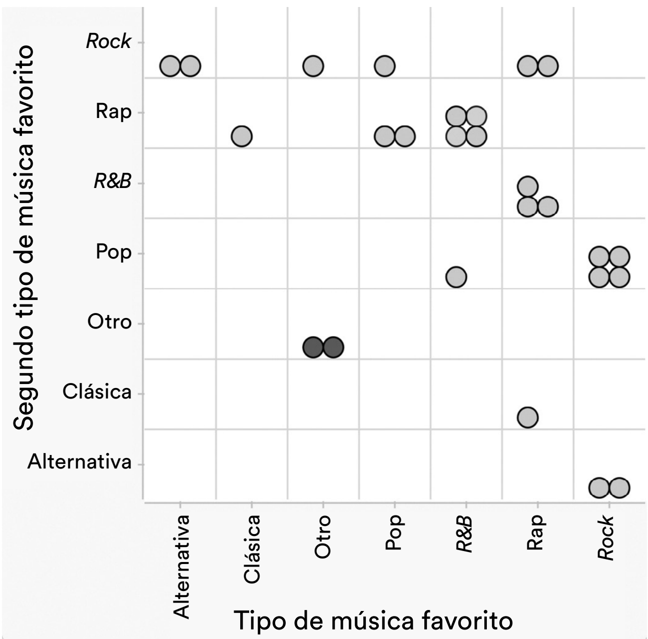


Figura 24. Primer y segundo tipo de música favoritos

segunda). La casilla de la esquina superior izquierda tiene dos puntos, lo que representa a dos estudiantes de la clase que han respondido que su música favorita es la alternativa y, en segundo lugar, el *rock*.

El análisis del primer y el segundo tipo de música favoritos muestra que casi todo el alumnado de la clase de 1.º de secundaria (24 personas, representadas por los puntos más claros en la figura 24) ha elegido *R&B*, rap o *rock* como su primera o segunda opción. Solo dos estudiantes (los puntos más oscuros) no incluyeron el *R&B*, el rap o el *rock* en ninguna de las dos primeras posiciones.

El alumnado también puede analizar la elección entre una banda en vivo y un *DJ*, y añadirlo a sus conclusiones finales sobre los tipos de música que le gustan (véase la figura 25). La diferencia entre el número de estudiantes que prefieren una banda en vivo frente a un *DJ* es cero.

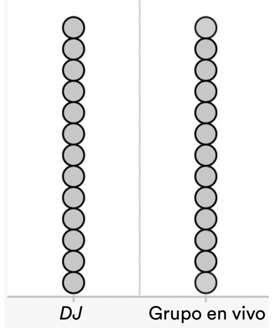


Figura 25. Elección entre grupo en vivo o *DJ*

Interpretar los resultados

El análisis muestra que el *R&B*, el rap y el *rock* son muy similares en cuanto al número de estudiantes que los eligieron como su tipo de música favorito. En total, 17 de los 26 estudiantes de la clase eligieron el *R&B*, el rap o el *rock* como su tipo de música favorito. Por otra parte, el alumnado del nivel B también podría señalar que los miembros de la clase eligieron por igual a la banda en vivo y al *DJ*.

Ejemplo 2: La elección de la música para la fiesta escolar (continuación). Comparación de grupos

Formular preguntas de investigación estadística

En el nivel B, el alumnado puede basarse en los datos descritos anteriormente y compararlos con los de otro grupo, que podría ser otra clase de otro curso. La pregunta de investigación estadística podría ser la siguiente:

¿Qué diferencias hay entre las clases respecto al tipo de música que les gusta a los estudiantes?

Recoger/considerar los datos

Esta pregunta de investigación estadística requiere que se recojan datos de una segunda clase (p. ej., una clase de Matemáticas de 2.º de secundaria); se pueden usar las mismas preguntas de encuesta.

Analizar los datos

Como el tamaño de los grupos puede ser diferente, para hacer comparaciones, deben resumirse los resultados con frecuencias relativas o porcentajes. En el nivel B se utiliza el razonamiento proporcional para resumir e interpretar los datos en términos de fracciones y porcentajes.

Los resultados de la pregunta 1 para las dos clases se resumen en la tabla 8 mediante la frecuencia (absoluta) y la frecuencia relativa (en porcentajes).

Las tablas muestran las diferencias entre las dos clases. Por ejemplo, mientras que a una mayoría de la clase de 1.º de secundaria (clase 1) le gusta el *rock* (69 %), a la mayoría de la clase de 2.º de secundaria (clase 2) no le gusta el *rock* (68 %). En el

Tabla 8. Frecuencias absolutas y relativas

Tipo de música	Clase 1 «Sí»		Clase 1 «No»		Total
Rap	20	77 %	6	23 %	26
Rock	18	69 %	8	31 %	26
Country	11	42 %	15	58 %	26
R&B	19	73 %	7	27 %	26
Pop	14	54 %	12	46 %	26
Clásica	11	42 %	15	58 %	26
Alternativa	14	54 %	12	46 %	26

Tipo de música	Clase 1 «Sí»		Clase 1 «No»		Total
Rap	20	77 %	6	23 %	26
Rock	18	69 %	8	31 %	26
Country	11	42 %	15	58 %	26
R&B	19	73 %	7	27 %	26
Pop	14	54 %	12	46 %	26
Clásica	11	42 %	15	58 %	26
Alternativa	14	54 %	12	46 %	26

gráfico de barras de la figura 26 se compara el porcentaje de estudiantes de cada una de las dos clases a quienes les gusta cada categoría musical.

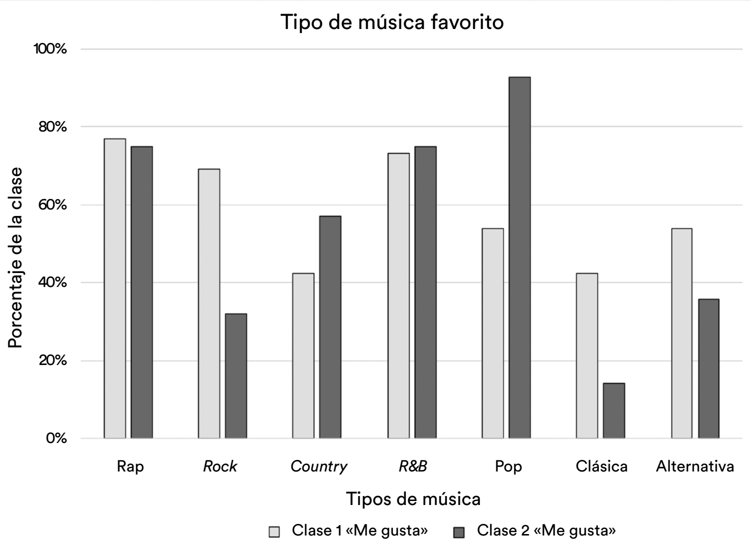


Figura 26. Gráfico de barras comparativo sobre las preferencias musicales

En el nivel B, el alumnado debe reconocer que no solo hay variabilidad de un individuo a otro dentro de un mismo grupo, sino que también hay variabilidad entre un grupo y otro. El segundo tipo de variabilidad lo ejemplifica el hecho de que el rap es el tipo de música más popular en la clase 1; mientras que, en la clase 2, lo es el pop. Es decir, la categoría modal en la clase 1 es el rap; y, en la clase 2, lo es el pop. A partir del gráfico de barras comparativo,[15] el alumnado debe darse cuenta de que el rap y el *R&B* tienen un nivel de popularidad similar en la clase 1 y en la clase 2; el *rock*, la música clásica y la alternativa son menos populares en la clase 2; y el *country* y el pop son más populares en la clase 2.

Tabla 9. Frecuencias para las clases 1 y 2 combinadas

Tipo de música	Clases 1 y 2 «Sí»		Clases 1 y 2 «No»		Total
Rap	41	76 %	13	24 %	54
Rock	27	50 %	27	50 %	54
Country	27	50 %	27	50 %	54
R&B	40	74 %	14	26 %	54
Pop	40	74 %	14	26 %	54
Clásica	15	28 %	39	72 %	54
Alternativa	24	44 %	30	56 %	54

Los resultados de cada clase también pueden combinarse para crear una muestra más grande (véase la tabla 9). Al combinar los datos entre clases, se observa que el tipo de música más popular es el rap, con 41 votos; seguido del *R&B* y del pop, con 40 votos cada uno.

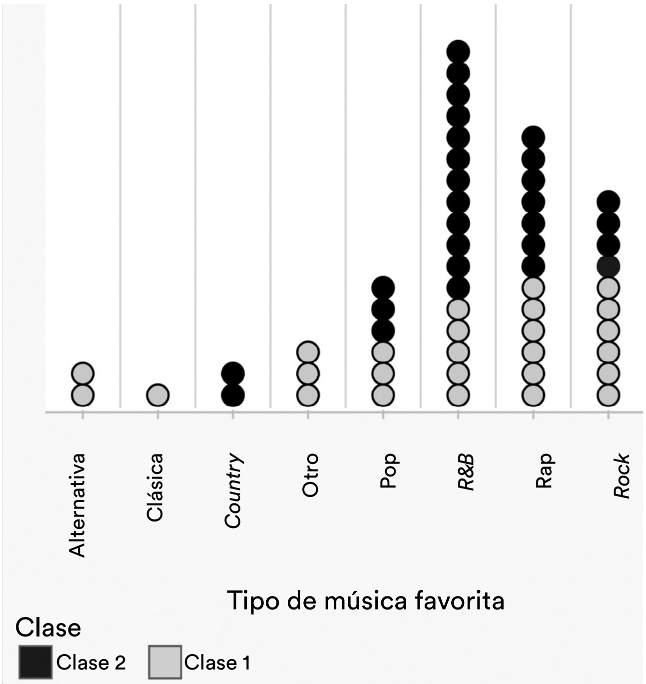


Figura 27. Tipo de música favorito conjunto para ambas clases

Además, el alumnado puede considerar los resultados de la segunda pregunta de la encuesta sobre el tipo de música favorito. En la figura 27 se combinan esos resultados para ambas clases. Debe darse cuenta también de que el tipo de música favorito es el *R&B* (31 %). La música clásica fue la elección menos popular como tipo de música favorito (2 %). El rap (24 %) y el *rock* (19 %) también destacan como preferidos por encima de otros tipos de música. Dado que es posible identificar de qué clase procede cada resultado (según leyenda de la figura 27), es fácil ver que la preferencia mayoritaria por el *R&B* de la clase 2 es lo que ha hecho que este tipo de música sea el favorito absoluto al combinar las dos clases.

[15] N. de T.: Los gráficos de barras comparativos pueden ser de varios tipos. En concreto, hay el gráfico de barras doble, que compara múltiples series de datos colocando barras de diferentes colores juntas para cada categoría (véase la figura 26); y el gráfico de barras apilado, que muestra la composición de un total dividiendo cada barra en segmentos que representan diferentes subgrupos (véase la figura 27).

Interpretar los resultados

El análisis muestra que las preferencias de tipos de música más comunes para el alumnado del instituto dependen de la clase. A la clase 1, de 1.º de secundaria, le gusta el rap, el *R&B* y el *rock.* A la clase 2, de 2.º de secundaria, le gusta el pop, el *R&B* y el rap. Los tipos de música favoritos para la clase 1 son el rap, el *rock* y el *R&B* (todos con porcentajes similares); para la clase 2, el tipo de música favorito es el *R&B*.

Dado que el *R&B* tiene el mayor número de respuestas como música favorita, tanto para la clase de 2.º de secundaria como para la combinación de ambas clases, el alumnado podría argumentar a favor del *R&B* como tipo de música para la fiesta. Sin embargo, puesto que más del 40 % de las personas encuestadas respondieron que preferían el rap o el *rock*, quizá sería mejor un *DJ* que pusiera una mezcla de los tres tipos de música favoritos.

En el nivel B, el alumnado debe reconocer que, aunque la muestra combinada sea mayor, es posible que aún no sea representativa de la población completa (es decir, de la totalidad del alumnado del centro educativo). En estadística, la aleatoriedad y la probabilidad se incorporan al procedimiento de selección de las muestras para proporcionar un método que sea «justo» y para mejorar las posibilidades de elegir una muestra representativa. Por ejemplo, si la clase decide tomar lo que se denomina una *muestra aleatoria simple* de 54 estudiantes, entonces cada posible muestra de 54 estudiantes tiene la misma probabilidad de ser elegida. Este uso ilustra una de las funciones de la probabilidad en la estadística. Aun siendo posible que el alumnado de este nivel no emplee realmente un proceso de selección aleatorio para recoger los datos, se deberían abordar las cuestiones relacionadas con la obtención de muestras representativas.

Ejemplo 3: La elección de la música para la fiesta escolar (continuación). Conexión entre dos variables cualitativas

Formular preguntas de investigación estadística

En el nivel B, el alumnado puede plantear preguntas de investigación estadística adicionales que conecten las dos variables cualitativas como:

A los estudiantes a quienes les gusta el rap, ¿suele gustarles o no gustarles el R&B?

¿Los estudiantes a quienes les gusta el rap tienen una mayor tendencia a que les guste el R&B en comparación con los estudiantes a quienes no les gusta el rap?

Estas preguntas requieren que el alumnado del nivel B condicione una de las variables (¿a los estudiantes les gusta el rap?) y considere las proporciones de estudiantes a quienes les gusta el *R&B* en las dos categorías de respuesta sobre el rap («Sí»/«No»). La primera pregunta solo tiene en cuenta la categoría «Sí» para determinar si a estos estudiantes les gusta el *R&B* más de lo que les disgusta. En cambio, la segunda pregunta comienza a evaluar la asociación entre dos variables cualitativas y compara las proporciones para las categorías «Sí» y «No».

Recoger/considerar los datos

El alumnado puede usar los datos recogidos en ambas clases para responder esta pregunta de investigación estadística. La pregunta de encuesta 1 pedía a cada estudiante si le gustaban o no diferentes tipos de música.

Analizar los datos

Las dos variables cualitativas pueden resumirse en una tabla de doble entrada (también llamada *tabla de contingencia*), que proporciona una forma de investigar posibles conexiones entre dos variables cualitativas. En este tipo de tablas, los datos suelen resumirse mediante recuentos (frecuencias). En

Tabla 10. Tabla de doble entrada de los «Sí»/«No» para el rap y el *R&B*

		Rap		
		Sí	**No**	**Total**
R&B	**Sí**	36	4	40
	No	5	9	14
	Total	41	13	54

este ejemplo, las categorías son si a los estudiantes les gustan («Sí») o no les gustan («No») el rap y el *R&B* (véase la tabla 10).

Otra opción es elaborar un gráfico de doble entrada para ilustrar los datos. En la figura 28, cada punto representa a una persona de la clase que contestó la encuesta.

Este gráfico de doble entrada muestra cuatro celdas (dos posibles elecciones para el rap: sí o no, por dos posibles elecciones para el *R&B*: sí o no). La celda superior izquierda tiene 36 puntos, que representan a las 36 personas que han dicho que les gustan el rap y el *R&B*. En este nivel, el alumnado debe ser capaz de describir la forma en que se organizan los datos en las tablas de doble entrada, así como de señalar los beneficios que puede proporcionar una representación visual.

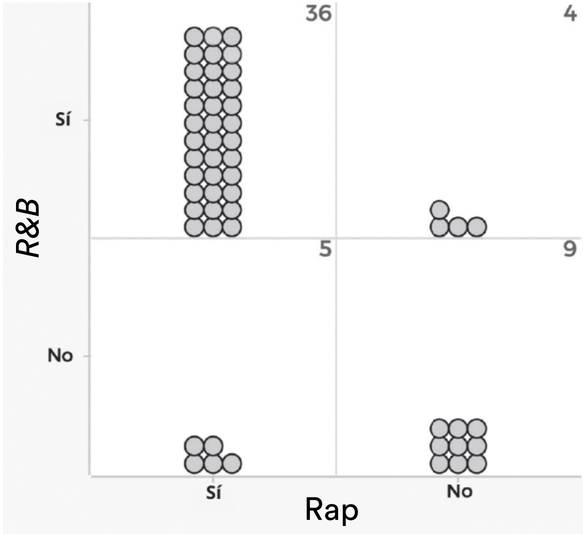


Figura 28. Gráfico de doble entrada con los recuentos

La tabla de contingencia y el gráfico de doble entrada muestran que, de las 41 personas a las que les gusta el rap, a 36 también les gusta el *R&B*. Es decir, a un 88 % (36/41) de los estudiantes a quienes les gusta el rap también les gusta el *R&B*. Por otro lado, el *R&B* no le gusta a un 12 % (5/41) de los estudiantes a quienes les gusta el rap. En cambio, si se fija la condición de que no les guste el rap, el *R&B* le gusta a solo un 31 % (4/13). El porcentaje de todas las personas a quienes les gusta el *R&B* es del 74 % (40/54). De los estudiantes a quienes les gusta el *R&B*, el porcentaje de personas a las que les gusta el rap (90 %) es mayor que el de las personas a las que no les gusta (10 %). El alumnado del nivel B debe prestar atención al uso del razonamiento proporcional en el análisis de datos cualitativos para dos variables.

El gráfico de mosaico también puede ser útil para analizar la posible relación entre variables cualitativas (véase la figura 29).

El gráfico de mosaico muestra visualmente que las proporciones de personas a las que les gusta el *R&B* en las dos categorías de rap son claramente diferentes. Las proporciones de los rectángulos más oscuros de las categorías del «Sí» y del «No» correspondientes al gusto por el rap son enormemente distintas. El rectángulo oscuro «Sí»/«Sí» ocupa el 88 % de la columna rap-«Sí», mientras que el rectángulo oscuro «Sí»/«No» solo ocupa el 31 % de la columna rap-«No».

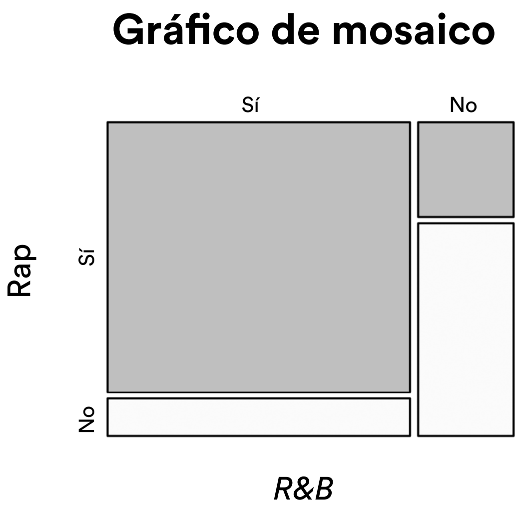


Figura 29. Gráfico de mosaico para las preferencias de *R&B* y rap

En el nivel B, el alumnado debe entender que las tablas de contingencia también pueden usarse para analizar probabilidades, especialmente probabilidades conjuntas y condicionales. Las celdas interiores proporcionan las frecuencias de los sucesos conjuntos que se necesitan para calcular la probabilidad de escoger al azar a alguien con una preferencia de tipo de música concreta en un grupo de 54 estudiantes. Por ejemplo, la probabilidad de elegir al azar a una persona a la que le gusten tanto el rap como el *R&B* es de 36/54. Las probabilidades condicionadas también utilizan las frecuencias conjuntas, pero consideran las frecuencias marginales en lugar del recuento total. Por ejemplo, la probabilidad de elegir al azar a una persona a la que le guste el *R&B* de entre los estudiantes a los que les gusta el rap es de 36/41. Si, en lugar de establecer la condición en el gusto por el rap, se condiciona en el otro sentido (el gusto por el *R&B*), se obtiene que la probabilidad de que a una persona le guste el rap, condicionada por que le guste el *R&B*, es de 36/40.

Interpretar los resultados

El análisis indica que las personas de esta muestra a las que les gusta el rap tienden a que les guste también el *R&B*. De las personas a las que les gusta el rap, a un 88 % les gusta el *R&B*, mientras que a un 12 % no. Además, quienes responden que les gusta el rap tienen una mayor tendencia a que les guste el *R&B* en comparación con quienes responden que no les gusta el rap. Entre las personas a las que les gusta el rap, la proporción de aquellas a las que les gusta el *R&B* es del 88 %; mientras que, entre las personas a las que no les gusta el rap, la proporción de aquellas a las que les gusta el *R&B* es solo del 31 %, una diferencia del 57 %. Esta diferencia sugiere que podría haber una asociación entre el rap y el *R&B*. Es importante que el alumnado del nivel B comprenda que existen limitaciones a la hora de generalizar estos resultados más allá de la clase.

Ejemplo 4: Los pinzones de Darwin. Comparación de una variable cuantitativa entre grupos

En el nivel B, el alumnado empezará a trabajar con datos que no provienen de su entorno inmediato. Por ejemplo, como parte de una investigación científica, pueden estudiar los pinzones de Darwin.

A lo largo de 2 millones de años, muchas especies de pinzones de las islas Galápagos han evolucionado a partir de un ancestro común: el semillero (la figura 30 muestra diez de esas especies). Estos pinzones han sido objeto de investigaciones científicas en todo el mundo desde que Darwin los observó por primera vez a principios del siglo XIX. Estas especies se han distribuido en diferentes espacios del ecosistema, donde pasan la mayor parte del tiempo y donde encuentran el tipo de alimentos que consumen.

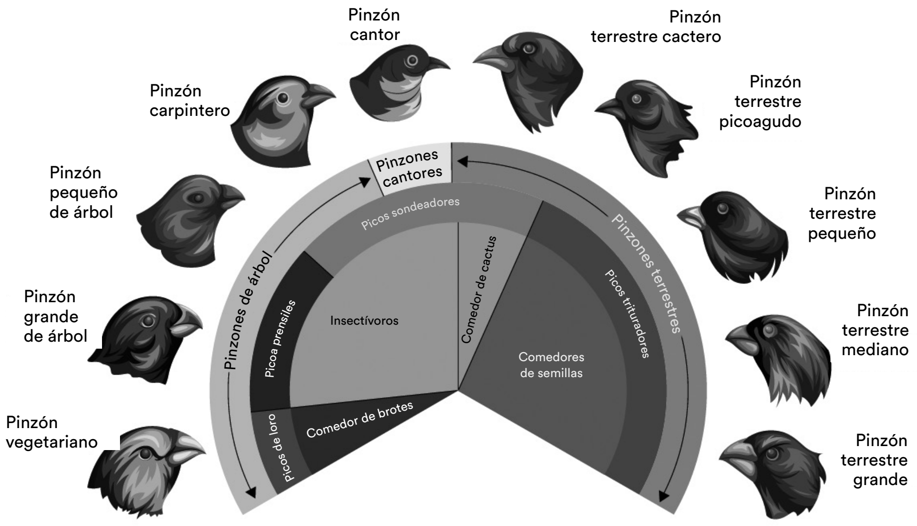


Figura 30. Especies de pinzones de las Galápagos. (Adaptada de Adobe Stock ##913646529.)

En el nivel B, el alumnado puede prever la variabilidad en los rasgos físicos (fenotipo) de una misma especie, tales como la altura, el peso, la longitud y la profundidad del pico. Esta variabilidad existe en la población debido a la variación de la expresión genética dentro de la especie.

Este ejemplo se centra en dos especies de pinzones terrestres en concreto: el pinzón cactero (*Geospiza scandens*) y el pinzón terrestre mediano (*Geospiza fortis*). El pinzón cactero se alimenta principalmente del fruto, las semillas, el néctar y el polen de la tuna, el cactus *Opuntia*. El pinzón terrestre mediano se considera generalista y se alimenta de muchos tipos diferentes de semillas. Sus distintos hábitos alimentarios pueden haber contribuido al desarrollo evolutivo de la forma de sus picos.

Formular preguntas de investigación estadística

El alumnado del nivel B puede analizar la variación dentro de una misma especie y entre distintas especies; y, en concreto, examinar las formas de los picos teniendo en cuenta su longitud y profundidad (véase la figura 31).

Figura 31. Profundidad del pico (mm) y longitud del pico (mm)

Los picos son una característica fundamental de las aves, que determinan su capacidad para acceder a determinados tipos de alimentos y, por tanto, tienen un impacto en su adaptación biológica. Se debe guiar al alumnado para que plantee preguntas de investigación estadística precisas e interesantes sobre la forma del pico de los pinzones, como:

Entre el pinzón cactero y el pinzón terrestre mediano de las islas Galápagos, ¿cuál de ellos suele tener el pico más largo?

El alumnado del nivel B empieza a comprender qué caracteriza a una buena pregunta de investigación estadística. Esta pregunta establece claramente la variable de interés (la longitud del pico) y las poblaciones de interés (los pinzones cacteros y los pinzones terrestres medianos de las islas Galápagos). Además, la intención de la pregunta es clara (comparar grupos), la pregunta puede responderse con los datos disponibles (existe información sobre las especies y sus longitudes del pico en el conjunto de datos) y, además, es una pregunta interesante y valiosa (ayudará a identificar posibles diferencias entre las dos especies de pinzones). Todas estas características hacen de esta una buena pregunta de investigación estadística. En www.nctm.org/gaise se incluye información detallada sobre la formulación de preguntas a lo largo del proceso de resolución de problemas estadísticos.

Recoger/considerar los datos

En el nivel B se debe aprender la diferencia entre datos primarios y secundarios, y empezar a comprender cómo examinar en detalle los datos secundarios y determinar cómo se recogieron estos datos, cómo se midieron las variables, cuáles son las unidades de observación y qué posibles resultados tienen esas variables. Para ayudar a responder la pregunta de investigación estadística sobre los pinzones, se pueden usar datos secundarios. El conjunto de datos es una muestra representativa de un conjunto de datos más grande (Grant y Grant, 2013) y contiene datos de 59 pinzones terrestres medianos y 42 pinzones cacteros. Las variables incluidas en los datos son la longitud del pico, la profundidad del pico, el año de recogida de los datos y si los datos se recogieron antes o después de una gran sequía. Nótese que, aunque el conjunto de datos se creó para tener alrededor de 30 observaciones por especie antes de la sequía y otras 30 después de la sequía, hubo menos datos disponibles del pinzón cactero después de la sequía.

Puesto que los datos son secundarios, el alumnado debería comenzar examinándolos. Atendiendo a la tabla de resumen (véase la tabla 11), puede identificar las variables incluidas en el conjunto de datos.

Tabla 11. Tabla de resumen de las variables disponibles

101 pinzones en total	
Nombre de las variables	**Descripción**
Anilla	N.º del pinzón
Especie	PC o PTM
Longitud del pico	Longitud del pico (mm)
Profundidad del pico	Profundidad del pico (mm)
Año	Año en que se registró la observación
Sequía	Antes o después de la sequía de 1977

La anilla representa el número de identificación de cada pinzón. La especie es una variable cualitativa con dos categorías: pinzón cactero (PC) y pinzón terrestre mediano (PTM). La longitud del pico es una variable cuantitativa continua que representa la medida de la longitud desde la base del pico hasta la punta, medida en milímetros (mm). La profundidad del pico es una variable cuantitativa continua que representa la altura de la parte más amplia del pico, también medida en milímetros.

La variable «Año» es una variable cuantitativa discreta que identifica el año en que se hicieron las observaciones. Por último, la variable «Sequía» es una variable cualitativa con dos categorías que indican si las observaciones fueron registradas antes o después de 1977, año en que hubo una gran sequía. El alumnado del nivel B debe comprender la diferencia entre variables cualitativas y cuantitativas.

En el nivel B se debe entender que, en estadística, con frecuencia, los resultados se generalizan del grupo estudiado a un grupo más grande: la población. En este ejemplo, el alumnado intenta obtener información sobre la población examinando una parte de esa población, llamada *muestra*. Estas generalizaciones son válidas solo si la muestra es representativa de ese grupo más grande. A medida que se progresa al nivel C, se desarrolla y se analiza el papel de la aleatoriedad como ayuda a la hora de obtener muestras representativas. Una muestra representativa es aquella en que las características relevantes de sus miembros son más o menos las mismas que las de los miembros de la población. Una selección inadecuada o sesgada de la muestra tiende a favorecer sistemáticamente determinados resultados y puede producir conclusiones engañosas y erróneas.

Analizar los datos

Para responder la pregunta de investigación estadística, se dibuja un gráfico de puntos comparativo inicial que muestra la longitud del pico según la especie (véase la figura 32).

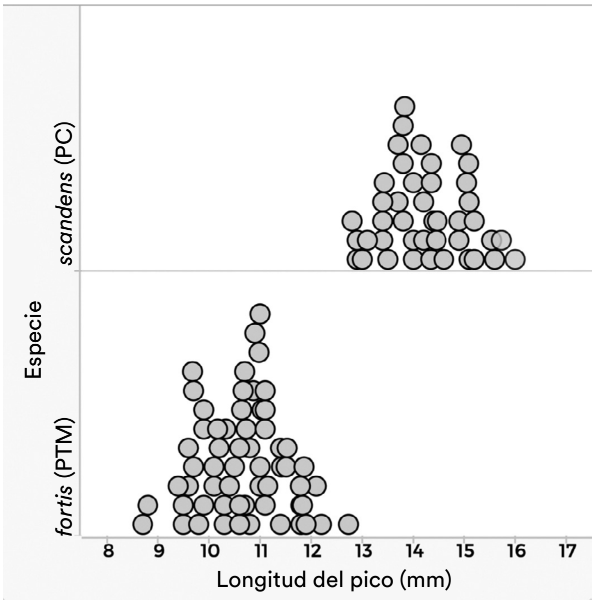


Figura 32. Gráficos de puntos comparativos para la longitud del pico

Este gráfico muestra que las longitudes del pico de la muestra de pinzones terrestres medianos (PTM) son aproximadamente simétricas y unimodales con el centro justo en los 11 mm. Las longitudes del pico de la muestra de pinzones cacteros (PC) también son aproximadamente simétricas y unimodales, con el centro ligeramente superior a los 14 mm. La longitud del pico de los PTM varía de los 8,7 mm a los 12,73 mm; y la longitud del pico de los PC, de los 12,8 mm a los 16 mm (véase la tabla 12).

Los problemas que requieren comparar distribuciones de dos o más grupos son habituales en estadística. Por ejemplo, en el nivel A, el alumnado podría comparar, mediante la comparación de distribuciones, el número de manchas que tienen las mariquitas de diferentes colores o la altura de las plantas de judías cultivadas en condiciones distintas. Uno de los gráficos más útiles para la comparación de datos cuantitativos entre dos grupos es el diagrama de caja.

El diagrama de caja (también llamado *diagrama de caja y bigotes*) es un gráfico basado en una división de los datos ordenados en cuatro partes con el mismo número de valores de datos en cada parte (aproximadamente, un cuarto de los datos). Los cuatro cuartos quedan determinados mediante el resumen de cinco valores: el valor mínimo de los datos («Mín.»), el primer cuartil («Q_1»), la mediana, el tercer cuartil («Q_3») y el valor máximo de los datos («Máx.»). Los resúmenes de estos cinco valores y los diagramas de caja comparativos para los datos de los pinzones se recogen en la tabla 12 y en la figura 33.

Tabla 12. Resumen de cinco valores correspondientes a la longitud del pico del PTM (*fortis*) y el PC (*scandens*); todas las medidas en milímetros

Índice	Especie	Mín.	Q_1	Mediana	Q_3	Máx.
1	*scandens* (PC)	12,8	13,7	14,2	14,94	16
2	*fortis* (PTM)	8,7	10,1	10,7	11,13	12,73

Para interpretar los diagramas de caja, el alumnado debe comparar las características globales de cada distribución (p. ej., centro y variabilidad alrededor del centro).

Por ejemplo, la mediana de la longitud del pico del PC es 14,2 mm, que es 3,5 mm mayor que la mediana del pico del PTM. La mediana sugiere que el valor típico de la longitud del pico del PC tiende a ser mayor que el del PTM.

Tanto la media como la mediana son medidas de tendencia central.[16] En el nivel A se presentó la mediana como la cantidad que tiene el mismo número de valores ordenados a cada lado. Esta «igualdad a cada lado» es lo que hace que la mediana sea una medida de tendencia central. La media también tiene interpretaciones diferentes en los niveles A y B. En el nivel A se presenta como el reparto equitativo, y en el nivel B se usa como punto de equilibrio (véase www.nctm.org/gaise).

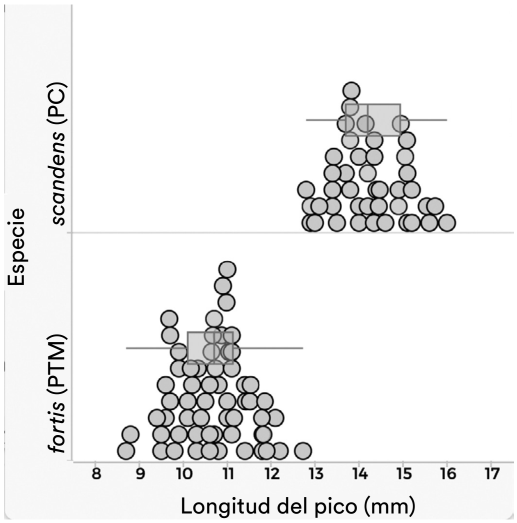


Figura 33. Gráficos de puntos comparativos con diagramas de caja para la longitud del pico

El rango para la longitud del pico del PTM es 4,03 mm, frente a 3,2 mm para el del PC. Los rangos indican que, en general, la distribución de longitudes del pico está más dispersa para el PTM que para el PC.

Otra medida de variabilidad que debería presentarse en el nivel B es el rango intercuartílico[17] (RIC). El RIC es la diferencia entre los cuartiles tercero y primero; indica el rango del 50 % central de los datos. El RIC para la longitud del pico es 1,03 mm en el caso del PTM y 1,24 en el del PC. Estos RIC muestran que la variabilidad en la mitad central de la distribución del PTM es ligeramente menor que la del PC.

La mediana y el 50 % central del diagrama de caja se pueden usar de manera conjunta para determinar visualmente si la longitud del pico del PTM es mayor o menor que la del PC en las respectivas poblaciones. Considerando que el 50 % central de los datos del PC no se solapa con el 50 % central de los datos del PTM, sería razonable afirmar que, en la población general de pinzones, la longitud del pico de los PC tiende a ser mayor que la de los PTM.

Interpretar los resultados

Dado que no hay superposición entre los diagramas de caja, estos datos sugieren que, en las islas Galápagos, los pinzones cacteros tienden a tener picos más largos que los pinzones terrestres medianos. El alumnado debe darse cuenta de que, si tomara otra muestra aleatoria, sería probable que obtuviera conclusiones similares.

En el nivel B también se debe comprender la noción de la variabilidad muestral, es decir, cómo varían los valores de un estadístico (como la mediana, la media, el rango, el RIC, etc.) de una muestra a otra. Por ejemplo, alguien podría tomar al mismo tiempo otra muestra de pinzones terrestres medianos en las Galápagos y obtener una mediana de la longitud de sus picos diferente. En el nivel C se desarrollarán herramientas que proporcionen un cierto grado de confianza para realizar la inferencia desde una muestra a la población.

Ejemplo 5: Los pinzones de Darwin (continuación). Separación frente a solapamiento

Formular preguntas de investigación estadística

El alumnado puede explorar más a fondo el conjunto de datos de los pinzones utilizando el proceso de resolución de problemas estadísticos para responder la pregunta de investigación estadística.

[16] N. de T.: Las medidas de tendencia central también son conocidas como *medidas de centralización*.
[17] N. de T.: También se denomina *recorrido intercuartílico*.

Entre el pinzón cactero y el pinzón terrestre mediano de las islas Galápagos, ¿cuál de ellos suele tener el pico más largo?

El profesorado puede animar al alumnado a hacer predicciones sobre lo que esperan ver al responder esta pregunta de investigación estadística. Con el conocimiento previo sobre la longitud del pico, al alumnado puede formular una conjetura fundamentada.

Recoger/considerar los datos

Se puede utilizar el mismo conjunto de datos descrito en el ejemplo anterior.

Analizar los datos

Esta pregunta de investigación estadística conduce a un análisis similar de los datos utilizando gráficos de puntos y de caja comparativos; sin embargo, los resultados no son tan claros como en el análisis anterior. El alumnado del nivel B debe tener la oportunidad de enfrentarse a datos que requieren un análisis con más matices. Asimismo, tiene que reflexionar sobre las limitaciones de los datos considerados y de los análisis realizados.

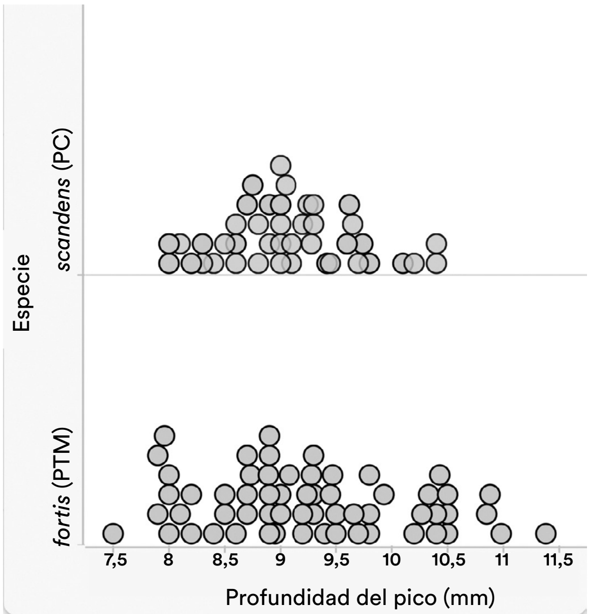


Figura 34. Gráficos de puntos comparativos para la profundidad del pico

El gráfico inicial de puntos comparativo de la profundidad del pico, clasificado por especies, se encuentra en la figura 34.

El gráfico muestra que las profundidades del pico para la muestra de pinzones terrestres medianos (PTM) son aproximadamente simétricas y unimodales con el centro un poco por encima de los 9 mm. Las de la muestra de pinzones cacteros (PC) son aproximadamente simétricas y unimodales con el centro por encima de los 9 mm. En el nivel B se debe entender que el punto más alto de la distribución puede no ser su centro. Las profundidades del pico para el PTM varían de los 7,5 mm a los 11,38 mm, y las del PC varían de los 8 mm a los 10,4 mm (véase la tabla 13).

Tabla 13. Resumen de cinco valores para las profundidades del pico

Índice	Especie	Mín.	Q_1	Mediana	Q_3	Máx.
1	*scandens* (PC)	8	8,63	9	9,44	10,4
2	*fortis* (PTM)	7,5	8,7	9,2	9,8	11,38

Un enfoque que funciona bien con distribuciones simétricas es describirlas utilizando la media como medida de tendencia central y, después, desarrollar una forma de describir la variabilidad a partir de la media.

Dado que las distribuciones no están sesgadas ni son bimodales, en este caso, la media es una medida de tendencia central apropiada. La profundidad media del pico para el PTM es 9,26 mm y para el PC es 9,07 mm.

La variabilidad puede describirse mediante el rango y el rango intercuartílico. También se puede describir comprobando cuánto se alejan los datos de la media. Esta medida de la variabilidad se denomina *desviación absoluta media* (DAM). La DAM es la distancia promedio de los datos a la media. Es decir:

$$\text{DAM} = \frac{\text{Distancia total a la media del conjunto de los valores}}{\text{Número de valores}}$$

La DAM para el PTM es 0,73 mm y para el PC es 0,49 mm. Esto indica que las profundidades del pico individuales para el PTM difieren de la media de 9,26 mm un promedio de 0,73 mm. Las profundidades del pico individuales para el PC difieren de la media de 9,07 mm un promedio de 0,49 mm. La DAM también sirve como un precursor de la desviación típica,[18] que se desarrolla en el nivel C.

En www.nctm.org/gaise se incluye una investigación detallada que explora la media y la DAM.

Siguiendo en la misma línea de la primera pregunta de investigación estadística, aquí también se han realizado los diagramas de caja comparativos. La figura 35 muestra que hay mucho solapamiento entre ambos diagramas de caja.

El rango del 50 % central —que mide el RIC— de las profundidades del pico para el PTM es 1,1 mm, con profundidades que varían de los 8,7 mm a los 9,8 mm; y el rango del 50 % central para el PC es 0,81 mm, con profundidades que varían de los 8,63 mm a los 9,44 mm. Las cajas de los diagramas de caja con los datos muestrales se solapan casi por completo, y ambas medianas se encuentran en la zona solapada.

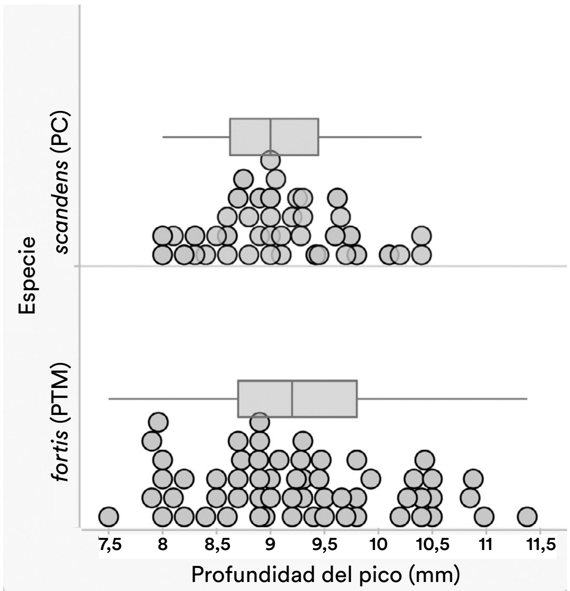


Figura 35. Diagramas de caja comparativos para la profundidad del pico del pinzón cactero (*scandens*) y el pinzón terrestre mediano (*fortis*)

Interpretar los resultados

No hay una indicación clara de si la profundidad del pico de los pinzones terrestres medianos de las islas Galápagos es mayor o menor que la de los pinzones cacteros. La profundidad media del pico difiere en 0,19 mm, lo que es relativamente poco para los datos proporcionados.

Cuando los resultados no son inmediatamente evidentes, el alumnado del nivel B debe reconocer la ambigüedad, teniendo en cuenta la variación entre muestras. Si se han tomado repetidas muestras de los pinzones, se puede hallar la profundidad media del pico para cada una. Se espera que haya cierta variabilidad en estas muestras. Una muestra diferente podría tener una profundidad media del pico para el PC mayor que la del PTM.

Incorporar la aleatoriedad en el proceso de muestreo permite usar la probabilidad para describir el comportamiento a largo plazo en la variabilidad de las medias de distintas muestras. Supongamos que cada estudiante toma una muestra aleatoria de pinzones y calcula la media muestral. Entonces, se podrían compilar las medias muestrales de cada estudiante para crear una distribución de las medias muestrales de la clase. El alumnado también puede empezar a hacer afirmaciones probabilísticas informales sobre qué valores de la media muestral son más probables y cuáles menos, observando cómo varían sus medias muestrales del valor más bajo al más alto de la clase. La variación en los resultados de muestreos aleatorios reiterados se describe a través de lo que se denomina *distribución muestral*. El comportamiento de esta distribución muestral permite cuantificar cómo de cerca se espera que esté una media muestral de la verdadera media poblacional. Las distribuciones muestrales se exploran más a fondo en el nivel C. Los recursos en línea disponibles en www.nctm.org/gaise también incluyen una tarea introductoria de investigación que explora el muestreo aleatorio, las distribuciones muestrales y la variabilidad muestral a través de simulaciones apropiadas para la transición del nivel B al nivel C.

[18] N. de T.: La desviación típica también se denomina en la literatura *desviación estándar*. Son términos sinónimos.

Ejemplo 6: Los pinzones de Darwin (continuación). La medida de la intensidad de la asociación entre dos variables cuantitativas

En el nivel B deben desarrollarse representaciones de datos más sofisticadas para investigar problemas que impliquen la relación entre dos variables cuantitativas (numéricas). El ejemplo de los pinzones puede extenderse para este propósito.

Formular preguntas de investigación estadística

Una observación interesante que el alumnado podría hacer ante las imágenes de los pinzones es que la longitud y la profundidad del pico podrían estar asociadas de alguna forma. Una pregunta de investigación estadística podría ser:

¿Existe una asociación entre la longitud y la profundidad del pico del pinzón terrestre mediano?

Recoger/considerar los datos

Para responder esta pregunta de investigación estadística, puede usarse el mismo conjunto de datos descrito en los ejemplos anteriores.

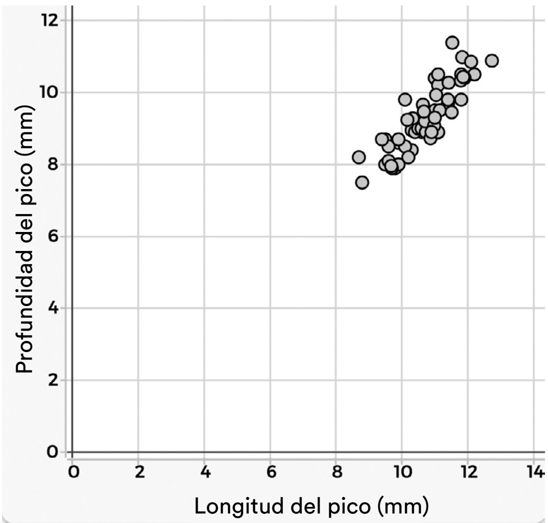


Figura 36a. Longitud y profundidad del pico para el pinzón terrestre mediano con origen en cero en ambos ejes

Analizar los datos

Para analizar la asociación entre dos variables cuantitativas, el alumnado del nivel B debe construir un diagrama de dispersión. Se puede hacer a mano con conjuntos pequeños de datos o mediante la tecnología para conjuntos mayores de datos. La longitud y la profundidad del pico para los pinzones terrestres medianos están representadas en las figuras 36a y 36b. Como en el nivel A, el alumnado del nivel B debe darse cuenta de cuándo los ejes horizontal y vertical no empiezan en cero y de cómo los incrementos en estos ejes pueden tener un impacto en el patrón visual. Por ejemplo, se puede comparar la figura 36a, donde los ejes empiezan en el origen, con la figura 36b, donde los ejes empiezan en el valor predeterminado del *software*. El diagrama de dispersión de la figura 36b sugiere una relación positiva entre la longitud y la profundidad del pico; a medida que aumenta la longitud del pico, también lo hace su profundidad. Además, la relación parece ser bastante lineal.

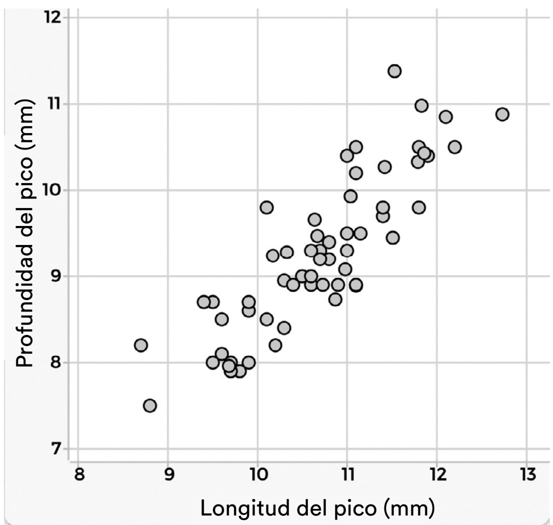


Figura 36b. Longitud y profundidad del pico para el pinzón terrestre mediano sin origen en cero

La medida de la intensidad de la asociación entre dos variables es un concepto estadístico importante que debería introducirse en este nivel B. Este concepto se desarrolla más a fondo en el nivel C y en niveles posteriores; sin embargo, la razón de recuento de cuadrantes (RRC) puede utilizarse ya en el nivel B para medir correlaciones.

Para ayudar al alumnado a identificar visualmente patrones en la distribución de puntos de un diagrama de dispersión, lo dividimos en cuatro regiones (o cuadrantes) mediante dos rectas. El diagrama de la figura 37, correspondiente a los datos de la longitud y la profundidad del pico, incluye una recta vertical trazada en la mediana de la longitud ($x = 10{,}7$ mm) y una recta horizontal en la mediana de la profundidad ($y = 9{,}2$ mm). Nótese que estas rectas también pueden trazarse usando las medias (véase en www.nctm.org/gaise) en lugar de las medianas.

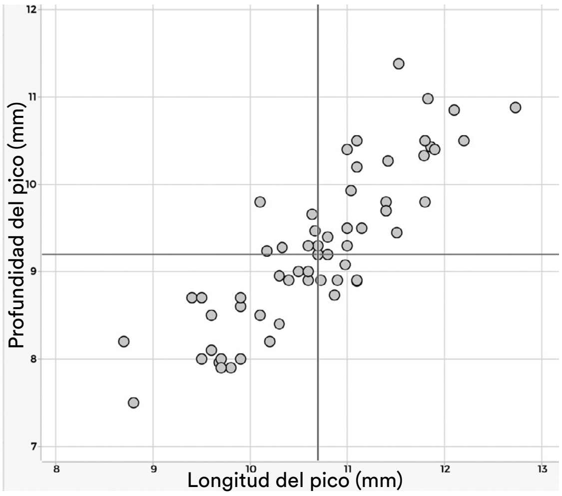


Figura 37. Diagrama de dispersión de la longitud y la profundidad del pico con las medianas indicadas por las rectas

La región superior derecha (cuadrante 1) contiene los puntos correspondientes a los pinzones con longitudes y profundidades del pico mayores que las medianas. La región superior izquierda (cuadrante 2) contiene los puntos que corresponden a los pinzones con longitudes del pico inferiores a la mediana y profundidades del pico mayores que la mediana. La región inferior izquierda (cuadrante 3) contiene los puntos correspondientes a los pinzones con longitudes y profundidades del pico menores que las medianas. La región inferior derecha (cuadrante 4) contiene los puntos que corresponden a los pinzones con longitudes del pico mayores que la mediana y profundidades del pico menores que la mediana.

La mayoría de los puntos del diagrama de dispersión están en el cuadrante 1 o en el cuadrante 3. Es decir, la mayoría de PTM con longitudes del pico mayores que su mediana tienen también profundidades del pico mayores que su mediana (cuadrante 1), y la mayoría de PTM con longitudes del pico menores que su mediana tienen también profundidades del pico menores que su mediana (cuadrante 3). Estos resultados indican que hay una asociación positiva entre las dos variables.

En términos generales, dos variables cuantitativas están asociadas positivamente cuando los valores por encima de la media de una variable tienden a aparecer conjuntamente con valores por encima de la media de la otra variable, y viceversa. La asociación negativa entre dos variables cuantitativas se da cuando los valores por debajo de la media de una variable tienden a aparecer conjuntamente con los valores por encima de la media de la otra variable.

Una medida de correlación es una cantidad que mide la dirección y la intensidad de la relación entre dos variables cuantitativas. El alumnado del nivel B debe notar que los puntos de los cuadrantes 1 y 3 (en este caso, un total de 44 puntos) contribuyen a la relación positiva entre la longitud y la profundidad del pico. Los puntos de los cuadrantes 2 y 4 (un total de 10 puntos) contribuyen a la relación negativa entre estas dos variables.

Una medida de correlación entre longitud y profundidad del pico es la razón de recuento de cuadrantes (RRC):

$$RRC = \frac{(n_{C1} + n_{C3}) - (n_{C2} + n_{C4})}{n} = \frac{44 - 10}{54} = 0{,}63$$

La RRC se calcula tomando la diferencia entre el número de puntos que son consistentes con una correlación positiva y el número de puntos consistentes con una correlación negativa, y dividiéndola después entre el número total de puntos. Es adimensional y sus valores están siempre entre −1 y +1, incluyendo ambos extremos.

Se debe animar al alumnado del nivel B a buscar tendencias generales en un diagrama de dispersión. Si la tendencia es lineal, cada estudiante puede trazar sobre el diagrama la recta que considere que describe mejor esa tendencia. En este nivel, esto debería hacerse de manera exploratoria. Dado que en otras áreas del currículo matemático estudian relaciones lineales, es una buena oportunidad para conectar el álgebra con la estadística. El grado de desarrollo que hayan alcanzado estas ideas determina cómo las puede explorar el alumnado.

Interpretar los resultados

El diagrama de dispersión muestra una relación lineal positiva entre la longitud y la profundidad del pico. La RRC de 0,63 también muestra que existe una relación positiva relativamente fuerte entre las dos variables. Estos resultados sugieren que la longitud del pico del PTM está positivamente relacionada con la profundidad de su pico. En el nivel C, la RRC se usa como base para el desarrollo del coeficiente de correlación de Pearson. En ese nivel también se presentan maneras más formales de explorar las relaciones lineales entre dos variables. Kader y Franklin (2008) abordan con más detalle las ventajas y las limitaciones de la RRC como medida de correlación. En www.nctm.org/gaise se recogen medidas de relación para variables cualitativas.

Ejemplo 7: Los pinzones de Darwin (continuación). Series temporales

Otro tipo importante de preguntas de investigación estadística que deben considerarse en el nivel B son las que requieren explorar tendencias en los datos a lo largo del tiempo. Los datos para la investigación de estas tendencias se conocen como *series temporales* y son bastante comunes en la ciencia.

Formular preguntas de investigación estadística

El ejemplo de los pinzones puede extenderse a una pregunta de investigación estadística relativa al tiempo:

¿Cómo cambia la longitud media de los pinzones terrestres medianos a lo largo del tiempo?

Recoger/considerar los datos

En el conjunto de datos general sobre los pinzones hay subconjuntos de datos que pueden usarse para responder esta pregunta de investigación estadística (Grant y Grant, 2013, figs. 01-06 [también 7.3]).

La figura 38 muestra que las dos variables necesarias para responder la pregunta de investigación estadística (año y longitud media del pico) están disponibles en el conjunto de datos.

Año	1973-2012
Especie	*fortis*
Longitud media del pico	10,28-11,11 mm
Profundidad media del pico	8,51-9,81 mm
Ancho medio del pico	8,27-8,93 mm
IC Longitud del pico	0,05-0,22
IC Profundidad del pico	0,05-0,22
IC Ancho del pico	0,04-0,17

Figura 38. Variables disponibles en el conjunto de datos

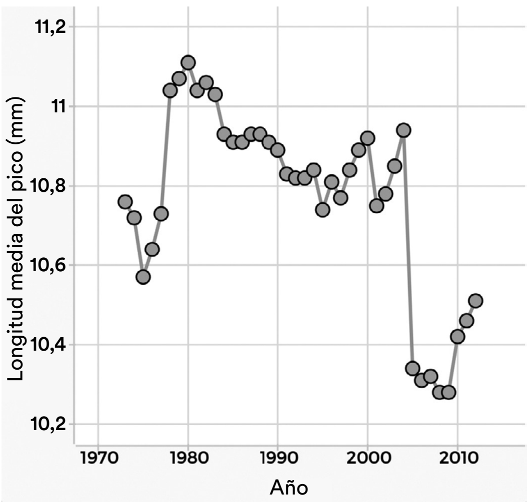


Figura 39. Gráfico de la serie temporal de la longitud media del pico para los pinzones terrestres medianos de 1973 a 2012

Analizar los datos

Para examinar el cambio en la longitud media del pico a lo largo del tiempo, se puede utilizar un gráfico temporal (véase la figura 39). El alumnado del nivel B debe razonar que el tiempo puede representarse en el eje horizontal; y la otra variable de interés, en el vertical. Siguiendo el espíritu de un diagrama de dispersión, el alumnado puede representar las longitudes registradas a lo largo del tiempo. Es apropiado conectar los puntos de este gráfico temporal para mostrar mejor las tendencias generales y los patrones a lo largo del tiempo.

Interpretar los resultados

Cualquier estudiante que examine el gráfico temporal puede darse cuenta de que las longitudes medias del pico para el PTM varían a lo largo del periodo de 35 años. Inicialmente, comienzan en 10,75 mm, disminuyen hasta 1975 y luego aumentan hasta alcanzar su nivel más alto en 1980. Después de 1980, muestran una tendencia general de disminución lenta, con una gran caída entre 2004 y 2005. Finalmente, la longitud del pico empezó a aumentar de nuevo en 2009. El alumnado puede investigar los pinzones terrestres medianos de las Galápagos para ver si encuentra explicaciones para las características de la serie temporal, en especial para los aumentos pronunciados y la gran caída. Asimismo, puede tratar de resolver preguntas como: ¿hubo un cambio en los alimentos disponibles esos años?, ¿el número de pinzones era el mismo en cada periodo de recogida de datos? En el nivel B puede describir lo que ve, y en el nivel C deberá progresar de la descripción de los datos a la realización de predicciones.

Ejemplo 8: Dollar Street. Imágenes como datos

Cuando el alumnado se adentra en el nivel B, debería empezar a comprender que los datos están por todas partes; se emplean para entender el mundo, incluyendo las cuestiones urgentes que son relevantes para la población mundial. Asimismo, debería comenzar a comprender que el proceso de resolución de problemas estadísticos le permite interactuar con conjuntos de datos grandes y complejos para interpretar mejor el mundo que le rodea. En este nivel, necesitará entender que el concepto *datos* también incluye imágenes, textos, sonidos, etc. Un ejercicio importante en este nivel B es familiarizarse con el manejo de diferentes tipos de datos, no solo de los almacenados en hojas de cálculo estáticas. Estos ejercicios se centran en que el alumnado trate de dar sentido a los datos no tradicionales.

Formular preguntas de investigación estadística

El uso de fotografías como datos, sumado al de los datos de entrevistas, ofrece una oportunidad para cuestionar las concepciones erróneas sobre familias, culturas y países, y para darse cuenta de que nuestras vidas cotidianas son con frecuencia más parecidas que diferentes. Anna Rosling Ronnlund (s.f.), la desarrolladora de Dollar Street, la página web interactiva de Gapminder, afirma que las personas de otras culturas son retratadas a menudo como aterradoras o exóticas, y que esto debe cambiar. Rosling quiere mostrar cómo vive realmente la gente, y le pareció natural usar fotografías como datos para que las personas puedan ver por sí mismas qué aspecto tiene la vida con diferentes niveles de ingresos. Dollar Street permite visitar muchos, muchos hogares de distintas partes del mundo, sin viajar. En el momento de preparar esta edición del informe en español, el proyecto Dollar Street dispone de imágenes de 469 familias, de 67 países distintos. En el conjunto de datos se incluyen más de 46 000 fotografías que pueden consultarse, junto con la información sobre las familias que viven en estos hogares, en Dollar Street (https://www.gapminder.org/dollar-street).

Utilizando los datos en forma de imágenes disponibles, el alumnado del nivel B podría analizar las respuestas a la siguiente pregunta de investigación estadística:

¿En qué se parecen o en qué se diferencian los conceptos familia *y* hogar *en distintas partes del mundo?*

Recoger/considerar los datos

Los datos en forma de imágenes son cada vez más comunes, ya que hacer fotografías es ahora más sencillo gracias a los móviles, y la gente toma fotos habitualmente de su día a día. El alumnado del nivel B puede trabajar con datos no convencionales, como los datos de imágenes, mediante la formulación de preguntas y la exploración. Para responder la pregunta de investigación estadística planteada, el alumnado de este nivel debe, primero, examinar en detalle los datos de Dollar Street, que son un conjunto de datos secundarios, para comprenderlos mejor. Mediante el examen de este tipo de datos, debe empezar a preguntarse y a investigar cómo y por qué se recogieron los datos, de quién se recogieron, cómo se midieron las variables y cuáles eran sus posibles resultados.

¿Cómo se recogieron los datos? Todas las familias que participan en Dollar Street lo hacen de forma voluntaria, conforme a unas directrices éticas adecuadas. Varias personas tomaron las fotografías, recogieron los

datos y pasaron un día en casa de cada familia, en los 67 países. En cada vivienda se tomaron imágenes de hasta 135 objetos, que después se etiquetaron y se añadieron al conjunto de datos. Además, se entrevistó a todas las familias exactamente con la misma batería de preguntas. El cuestionario utilizado añade el contexto necesario para las fotografías. Las familias que participan en Dollar Street son voluntarias, y se pueden añadir nuevas familias al conjunto de datos en cualquier momento. Para ello, deben firmar un consentimiento y completar un cuestionario sobre su contexto personal. El cuestionario se encuentra aquí:

https://docs.google.com/document/d/1vtracv6xSEDWvglYVDi7k2fAATs55wZUwnMdECIIfS4/edit#heading=h.x1v8rup1z5m0

¿De quién se recogieron los datos? Los datos se obtuvieron de familias de distintas partes del mundo. Una familia se considera una unidad de observación (un caso) en el conjunto de datos. Para cada familia se registraron numerosas variables, por lo que los datos son multivariantes.

Variables recogidas. Hay más de 150 variables definidas en el conjunto de datos, incluyendo camas, cepillos de dientes, puertas principales, áreas de trabajo, baños y paredes interiores. Las variables se registran a partir de las propias fotografías. Además de esas variables captadas por las imágenes, el conjunto de datos también contiene una variable que define una estimación del consumo mensual por persona adulta que reside en la vivienda, expresado en términos de dólares gastados. Lindgren (s.f.) expone las hipótesis y las limitaciones, y cómo se calcula el valor del consumo mensual.

Analizar los datos

El alumnado del nivel B debe disponer de tiempo suficiente para examinar estos datos. A menudo, con conjuntos de datos grandes o no tradicionales, necesita tiempo para digerirlos y, simplemente, «jugar» con ellos. Los datos son tan abundantes que resulta difícil analizar todas las variables y los matices de las imágenes si se dispone de poco tiempo. Además, gran parte del valor de trabajar con datos no tradicionales reside en perfeccionar la capacidad de hacer preguntas. Como comienzo de la investigación, se puede animar al alumnado a registrar varias observaciones y preguntas sobre las imágenes o la página web. Estas observaciones y preguntas le ayudarán a familiarizarse con el diseño de la página web y la complejidad del conjunto de datos, y permitirán al profesorado abordar las generalizaciones y los sesgos en los que el alumnado podría estar incurriendo involuntariamente. En este punto, es útil recordarle que el propósito del conjunto de datos de Dollar Street era cuantificar el consumo a través del análisis del dinero del que disponían las familias para sus gastos.

Transcurrido el tiempo para examinar los datos, se puede guiar al alumnado para que seleccione una variable, por ejemplo, la variable «Familias», como se ilustra en la figura 40.

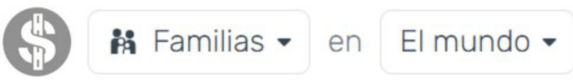


Figura 40. Variable «Familias»

Al elegir esta variable, se mostrarán imágenes de las diferentes familias que participan en Dollar Street (véase la figura 41). Se puede pedir al alumnado que ignore el texto que figura en las imágenes (importe del consumo mensual en dólares y ubicación de la foto), y que se centre únicamente en las imágenes. Pueden preguntarse:

¿Cómo son las fotografías típicas de una familia de Dollar Street?

El alumnado, en este nivel B, puede fijarse en que, en los distintos países del mundo, las personas que posan para las fotos están físicamente cerca, a menudo abrazándose. Los más pequeños están en brazos de un miembro de la familia. ¿Cómo puede


Figura 41. Familias de Dollar Street

describirse el lenguaje corporal que reflejan las fotografías? ¿Es similar en todo el mundo? Las familias se muestran relajadas; algunas personas sonríen, otras no. El alumnado también puede darse cuenta de que el tamaño de la familia (número de individuos identificados como familia) varía. Estas fotografías muestran que los grupos que se etiquetan como familia pueden ser diferentes. Algunas familias incluyen a una generación mayor (presumiblemente los abuelos y abuelas), otras tienen solo un progenitor y algunas tienen dos. Algunas familias no tienen descendencia y otras podrían tener hijos e hijas ya mayores que no aparecen en la imagen.

A pesar de las diferencias en el aspecto o la edad, la observación más notable es que el concepto de familia como un grupo unido parece ser el mismo en todos los países representados. Se puede animar al alumnado a hacer clic en una fotografía para conocer mejor a una familia en concreto.

Una segunda variable que se puede analizar en el nivel B es «Puertas principales», que indica la puerta de entrada a la vivienda de las familias (véase la figura 42).

Figura 42. Variable «Puertas principales»

El alumnado puede observar las imágenes de la puerta principal de la vivienda de las familias para describirlas, indicando los materiales típicos, el funcionamiento y el uso. En general, las fotografías de las puertas muestran algún tipo de elemento que cubre la entrada a la vivienda. En su forma más básica, una puerta es igual en todas partes. También parece que los materiales de las puertas son similares, mayoritariamente madera, aunque a veces es metal o tela. La figura 43 muestra cuatro puertas de distintos niveles económicos, todas hechas de madera.

Figura 43. Puertas de madera

Se observa cierta variación en las puertas cuando se considera todo el rango de consumo mensual. En el grupo de consumo bajo, por ejemplo, las puertas no suelen tener cerraduras y están menos decoradas; mientras que, en el grupo de consumo alto, todas las puertas tienen cerradura, a veces incluso varias.

Figura 44. Barra de selección

Para analizar la relación entre el nivel de consumo y los tipos de puertas, el alumnado puede categorizar el consumo en bajo, medio o alto. Por ejemplo, la categoría de consumo medio incluiría a las familias situadas en el rango medio de la escala «de más pobre a más rico» que figura en la página web. Para elegir solo a esas familias, se debe ajustar la barra de selección (véase la figura 44).

La figura 45 muestra algunas fotografías de puertas de las categorías de consumo: alto, medio y bajo.

Figura 45. Puertas correspondientes a diferentes consumos

Aunque hay diferencias entre las puertas, lo más llamativo son las similitudes. El alumnado puede apreciar que hay muy poca variación en las puertas de los distintos países, sobre todo si se concentra en una franja económica específica.

En el nivel B se pueden examinar también más variables, como la variable «Estufas»[19] (véase la figura 46).

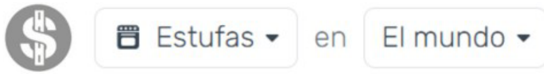


Figura 46. Variable «Estufas»

El alumnado puede observar varias diferencias en las fotografías de esta variable. Para empezar, las diferencias más notables se dan entre las familias que consumen menos mensualmente y las que consumen más (véase la figura 47).

Figura 47. «Estufas» para diferentes consumos

Las familias que consumen más al mes parecen tener cocinas eléctricas o de gas (véase la figura 48), integradas en una encimera o como electrodoméstico independiente. En las familias con un consumo más bajo, los fogones suelen estar en el suelo o consistir únicamente en un fuego. El uso de madera como combustible es habitual en las familias de consumo bajo-medio. Las cocinas de gas de las familias de consumo bajo suelen tener un único quemador, no varios. Aunque todas las familias del mundo usan calor para cocinar, la manera de hacerlo varía.

Figura 48. «Estufas» con distintas fuentes de calor

En una charla TED, la fundadora del proyecto Dollar Street, Anna Rosling Ronnlund (2017), aporta más información acerca del proyecto, así como la manera de cocinar en todo el mundo.

Interpretar los resultados

La pregunta de investigación estadística requería que el alumnado comparase familias, y sus viviendas, de distintas partes del mundo. En general, los datos de las imágenes revelaron un concepto de familia similar a lo largo de todo el espectro económico y en todos los países. Las puertas también eran similares, aunque había cierta variación en función del consumo. Por otra parte, los aparatos de cocina variaban según el consumo. En líneas generales, las viviendas parecían variar más económica que geográficamente.

El proyecto Dollar Street es, en este momento,[20] una colección de 469 familias. Ninguna de las conclusiones expuestas debería generalizarse, dado que se trata, más bien, de un estudio observacional de datos presentados como imágenes. El alumnado del nivel B debe comprender que las conclusiones extraídas de estos datos visuales tienen un alcance limitado debido a las propias limitaciones de los datos. El ejemplo de Dollar Street le permite tomar conciencia de cómo los datos fotográficos de las redes sociales pueden determinar o cambiar fácilmente percepciones o conclusiones que van mucho más allá del alcance real de esos datos. Desarrollar un escepticismo sano en términos estadísticos requiere plantear preguntas difíciles sobre los datos y sobre quienes los recogen, almacenan o presentan.

El alumnado del nivel B debe exponerse a datos no convencionales para ampliar su conceptualización del impacto de los datos y de la estadística en nuestro mundo. El objetivo es que el uso de conjuntos de datos como Dollar Street lleve al alumnado a reflexionar críticamente acerca de sus percepciones sobre cómo viven las personas en las distintas partes del mundo, y sobre la importancia de realizar más investigaciones sobre este tema.

[19] N. de T.: Mantenemos el término *estufa* porque es el utilizado en la página web de Dollar Street, como puede verse en la captura de pantalla de la figura 46. No obstante, a nuestro juicio, para el original *stoves* habría sido más acertada la traducción como «tipo de cocina» o «aparatos de cocina».

[20] N. de T.: Se refiere al momento de preparación de esta edición del GAISE II en español (2026).

Ejemplo 9: Memoria y música. Experimentos comparativos

Otro diseño de estudio importante y apropiado para abordar en el nivel B es el de experimentos comparativos. Estos estudios implican comparaciones de los efectos de dos o más tratamientos (condiciones experimentales) sobre una o varias variables de respuesta. En el nivel B es suficiente con estudios que comparen dos tratamientos. Este tipo de análisis lo lleva a cabo frecuentemente el alumnado que participa en competiciones científicas.

Formular preguntas de investigación estadística

Es habitual que el alumnado escuche música mientras estudia; a menudo aseguran que les ayuda a concentrarse y a tener un estado de ánimo positivo. Para investigar esta cuestión, el alumnado del nivel B puede diseñar un estudio comparativo experimental para analizar los efectos de escuchar música en la habilidad para memorizar palabras. Una pregunta de investigación estadística podría ser:

¿Los estudiantes son capaces de memorizar más palabras cuando escuchan música que cuando no la escuchan?

Recoger/considerar los datos

La clase debe desarrollar una estrategia de diseño para recoger los datos experimentales y responder la pregunta de investigación. Esto implicará que el alumnado identifique y, en la medida de lo posible, tenga en cuenta las potenciales fuentes de variabilidad no controladas que puedan interferir con la interpretación de los resultados.

Un experimento simple para realizar con una clase es dividirla aleatoriamente en dos grupos de igual tamaño. De este modo, el alumnado puede comprender que la asignación aleatoria es importante en el diseño experimental porque tiende a promediar las diferencias entre las habilidades de cada individuo, así como entre otras características que puedan afectar a la respuesta. Además, la asignación aleatoria permite establecer conjeturas causales.

Supongamos que se divide aleatoriamente a 28 estudiantes en dos grupos de 14. En la medida de lo posible, debe usarse la tecnología para llevar a cabo la asignación aleatoria a los grupos, aunque también pueden utilizarse otros recursos, como una baraja de cartas. En este caso, por ejemplo, la profesora baraja 28 cartas, 14 rojas y 14 negras, y entrega una a cada estudiante. Quienes reciben una carta roja constituyen el grupo que escuchará música; y quienes reciben una carta roja, el grupo que estará en silencio.

La clase debe elaborar un procedimiento experimental y ponerse de acuerdo sobre sus detalles. Por ejemplo, cada participante puede tener dos minutos para memorizar una lista de 20 palabras; hacer una pausa de dos minutos; y, después, disponer de otros dos minutos para escribir, de esas palabras, todas las posibles. Los miembros del grupo con música escucharán una determinada canción con letra durante el experimento, mientras que el grupo control permanecerá en silencio todo el rato. El número de palabras correctamente recordadas es la variable de respuesta de interés. El conjunto de datos, por lo tanto, incluye dos variables: el grupo al que se ha asignado aleatoriamente cada participante, que es una variable cualitativa; y el número de palabras correctamente recordadas, que es una variable cuantitativa discreta. El alumnado puede utilizar la tecnología para registrar estos datos.

Analizar los datos

La clase debe calcular los estadísticos de resumen correspondientes al grupo que escucha música y al que permanece en silencio. La tabla 14 muestra unos datos de ejemplo. En algún momento, el alumnado debería calcular manualmente estos números a partir de un conjunto de datos pequeño para comprender su composición. Sin embargo, con conjuntos mayores y cuando se recogen datos mediante la tecnología, es importante saber cómo usar esa tecnología para calcular el resumen de cinco medidas.

Los diagramas de caja con una misma escala para ambos grupos son una representación gráfica apropiada (véase la figura 49).

Interpretar los resultados

Estos resultados sugieren que los estudiantes que participan en el estudio memorizan, en general, menos palabras cuando escuchan música que cuando están en silencio. Con la excepción del valor máximo del grupo con música, el resumen de cinco medidas del grupo con música (M) son inferiores a las del grupo en silencio (S). La variación en el 50 % central de las puntuaciones es similar en ambos grupos. La distribución de S parece más o menos simétrica, mientras que la distribución de M está ligeramente sesgada a la derecha. Teniendo en cuenta la variación en las puntuaciones y la separación en los diagramas de caja (las medianas están fuera de la parte de las cajas centrales que se solapa), una diferencia de 3 entre las medianas es bastante grande.

Tabla 14. Resumen de cinco medidas correspondientes al número de palabras memorizadas en los grupos con música y en silencio

	Número de palabras memorizadas	
	Grupo con música	**Grupo en silencio**
Mínimo	3	6
Primer cuartil	6	8
Mediana	7	10
Tercer cuartil	9	12
Máximo	15	14

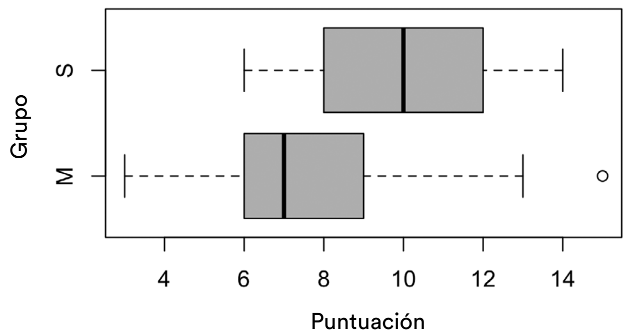


Figura 49. Diagramas de caja comparativos para los datos de memoria. M = música; S = silencio.

El alumnado puede preguntarse si el tipo de música fue un factor importante en los resultados. Por ejemplo, ¿contribuyó la presencia de palabras en la canción a la variación entre los diagramas de caja? Se puede valorar repetir el experimento con un nuevo grupo de estudiantes, esta vez usando solo música instrumental. En el nivel B se debe comprender el alcance de la inferencia de este experimento. Los estudiantes no fueron seleccionados aleatoriamente, aunque sí fueron asignados aleatoriamente a los tratamientos. La ausencia de un muestreo aleatorio limita el alcance de la inferencia a la clase, en vez de a la población general. El uso de una asignación aleatoria permite establecer conjeturas causales respecto a los tratamientos en la clase. En el nivel C se continuará profundizando en el papel de la aleatorización en estadística.

Resumen del nivel B

El alumnado que comprende los conceptos estadísticos del nivel B puede comenzar a valorar que el razonamiento estadístico es un proceso de resolución de problemas. Ganará experiencia en la formulación de sus propias preguntas de investigación estadística, en la recogida y la valoración de datos apropiados procedentes de distintas fuentes, en el análisis de datos mediante gráficos y medidas de resumen, y en la interpretación de resultados atendiendo a las posibles inferencias de causalidad o a las inferencias de una muestra a la población. A medida que el alumnado empieza a formular sus propias preguntas de investigación estadística, es más consciente de que el mundo que lo rodea está repleto de datos que afectan a su vida, y comienza a apreciar que la estadística puede ayudarle a tomar decisiones basadas en datos. Una vez comprendido el proceso de resolución de problemas estadísticos, se puede animar al alumnado del nivel B a que examine las afirmaciones de los medios de comunicación, como las relacionadas con las decisiones sanitarias (p. ej., las vacunas), los problemas medioambientales (el nivel del mar, los niveles de CO_2, etc.) o las políticas educativas (el currículo, la evaluación, etc.). Véase www.nctm.org/gaise para más ejemplos.

Cuando el alumnado se sienta cómodo con los conceptos del nivel B, estará preparado para la mayor profundización del nivel C.

Nivel C

Introducción

El papel de la tecnología

El papel de la probabilidad en estadística

Aspectos esenciales de cada componente

Ejemplo 1: Revisión del nivel B. Los pinzones de Darwin. De la desviación absoluta media a la desviación típica

Ejemplo 2: Revisión de los niveles A y B. La elección de la música para la fiesta escolar (continuación). Generalizar los resultados

Ejemplo 3: La elección de la música para la fiesta escolar (continuación). Inferencia sobre la asociación

Ejemplo 4: El efecto de la luz en el crecimiento de plántulas de rábano. Experimentos

Ejemplo 5: La consideración de las medidas en el diseño de prendas de vestir. Regresión lineal

Ejemplo 6: La siesta y los ataques al corazón. Inferir la asociación a partir de un estudio observacional

Ejemplo 7: La población en edad laboral. Trabajar con datos secundarios

Ejemplo 8: La clasificación de lagartijas. Predecir una variable cualitativa

Resumen del nivel C

Introducción

Las pautas del nivel C se basan en los fundamentos del razonamiento estadístico desarrollados en los niveles anteriores. En este nivel, las actividades ilustrativas deben seguir centrándose en los cuatro componentes del proceso de resolución de problemas estadísticos y afianzando el espíritu de la auténtica práctica estadística.

En el nivel C, el alumnado entiende cómo utilizar las preguntas a lo largo del proceso de resolución de problemas estadísticos y cómo usar la información procedente de los datos para responder de forma adecuada a las preguntas de investigación estadística formuladas. Asimismo, reconoce datos no tradicionales (textos, sonidos, imágenes) junto con datos tradicionales (números en contexto). Elige y usa las herramientas adecuadas para el análisis de datos (representaciones gráficas, representaciones tabulares y resúmenes numéricos). Además, en este nivel se desarrolla el pensamiento multivariante y se amplían los tipos de preguntas de investigación estadística para incluir preguntas relativas a la causalidad y la predicción.

En los niveles A y B se recopilaban los datos de grupos completos, muestras y experimentos simples. En el nivel C se profundiza en este tipo de estudios. El alumnado refuerza su comprensión del papel del muestreo y la asignación aleatoria. Mediante los estudios estadísticos del nivel C, desarrolla un entendimiento más formal del razonamiento inferencial y su uso de la probabilidad, y también se centra en explicar el razonamiento estadístico a otras personas.

Muchos de los datos con los que el alumnado se encuentra en su vida diaria no proceden de estudios formales, sino de sensores, los llamados *bots* (p. ej., el número de veces que hacemos clic en un anuncio concreto o el número de coches que pasa por un determinado cruce), o de interacciones con la tecnología (p. ej., dispositivos que registran el tiempo que pasamos a la semana frente a una pantalla). A veces, los datos pueden representar a toda la población, como los datos relacionados con las elecciones musicales de las personas que usan un servicio de *streaming* de música. El alumnado podría recoger datos mediante sus dispositivos móviles para documentar la incidencia de los grafitis en su ciudad o para identificar los pasos de cebra peligrosos. En el nivel C se debe comprender que estos datos no siempre son compatibles con un razonamiento estadístico inferencial formal. No obstante, estos datos pueden contar una historia parcial que describa posibles patrones o procesos, o que proporcione hipótesis para una recogida de datos más formal.

En el nivel C se desarrolla una noción de los datos más sofisticada, que tiene en cuenta los errores y los valores perdidos. Se profundiza en su análisis integrando la tecnología en su práctica, transformando las variables o creando otras nuevas. El alumnado entiende que los datos suelen almacenarse como archivos en un ordenador, en un servidor o en una nube; y que estos archivos pueden compartirse y modificarse, lo cual plantea la necesidad de prácticas reproducibles y del tratamiento ético de los datos.

El alumnado del nivel C va más allá de las estadísticas descriptivas de todo un grupo o población para incorporar las nociones de azar y probabilidad y realizar inferencias generales y comparaciones de tipo inferencial sobre dos grupos, según corresponda. Se utilizan simulaciones para mejorar el razonamiento probabilístico.

En este nivel se comprende la variabilidad entre muestras. Se pueden simular distribuciones muestrales aproximadas y usarlas para calcular p-valores. El p-valor[21] es una medida[22] de uso común para abordar la pregunta: ¿puede este resultado deberse únicamente al azar? En los últimos años se ha evaluado el uso y la interpretación del p-valor en los resultados de las investigaciones, lo que ha llevado a recomendar que se preste una mayor atención a la forma en la que se comunican los resultados estadísticos (Nuzzo, 2014; Wasserstein y Lazar, 2016). En estadística, el alumnado del nivel C debe entender la importancia de for-

---

[21] N. de T.: El p-valor (*p-value* en inglés) también se encuentra en la literatura estadística en español como *valor p* (también unido con guion), *nivel p* o *probabilidad de significación*.

[22] N. de T.: En el original se dice: «The p-value is a commonly used statistic», que puede ser válido en inglés. Sin embargo, como se señaló en una nota anterior, el término *estadístico* se suele reservar para una función de la muestra aleatoria. Dado que el p-valor es una probabilidad que se determina a partir de un estadístico, no es propiamente un estadístico. Para evitar la confusión con que pueda ser un valor concreto de un estadístico, tampoco consideramos adecuado referirnos a él como «estadística».

mular y responder la pregunta: ¿puede este resultado deberse únicamente al azar? Asimismo, el alumnado puede evaluar una distribución muestral simulada para determinar qué resultados se consideran extremos y qué resultados son probables.

El papel de la tecnología

Siempre que sea posible se debe usar la tecnología para desarrollar la comprensión de la estadística y abordar investigaciones en este campo. Integrar la tecnología en el currículo puede resultar un reto. No todas las aulas están equipadas con el *hardware* o el *software* necesarios, y los cambios tecnológicos son tan frecuentes que la planificación no siempre es fácil. La tecnología también puede agravar los problemas de equidad, dado que parte del alumnado tiene mayor acceso a ella o más apoyo tecnológico en casa que el resto. El apoyo tecnológico puede variar enormemente, incluso dentro de una misma zona. Aun así, la práctica estadística moderna no puede disociarse de la tecnología, y existen muchas herramientas de *software* y *applets*, disponibles gratuitamente, para mejorar la comprensión del alumnado, por lo que se recomienda adoptar la tecnología en la mayor medida posible.

Los conceptos estadísticos abstractos, como la noción de la variabilidad entre muestras, pueden hacerse más concretos gracias a la tecnología. En el mundo real, rara vez se dispone de oportunidades para obtener múltiples muestras aleatorias y observar directamente la variación de las estimaciones inducida por el azar. Sin embargo, las simulaciones y los *applets* permiten la observación directa de la variabilidad a partir de muestras repetidas y, desde ahí, se puede desarrollar una comprensión más profunda de la teoría. Mediante simulaciones y *applets* también es posible ilustrar la necesidad de la asignación aleatoria para realizar afirmaciones causales; de este modo se proporciona al alumnado una comprensión más profunda del papel del azar y la probabilidad en la inferencia estadística.

Con la inclusión y el fomento de la informática[23] en el currículo, cada vez está más presente el pensamiento computacional en los centros y en las políticas educativas. Una clase de Estadística en la que se usa un lenguaje de programación estadística aporta un complemento enriquecedor a las clases de informática. Por este motivo, la importancia del pensamiento estadístico es mayor que nunca. Puesto que los datos se usan en informática y otras disciplinas, resulta extremadamente importante que el alumnado del nivel C entienda el alcance de los datos y sus limitaciones, así como los tipos de preguntas que se puede hacer sobre ellos. También existen paquetes de *software* estadístico que han sido diseñados específicamente para estudiantes de estadística y que pueden facilitar un pensamiento estadístico más profundo. El profesorado debe elegir las tecnologías de acuerdo con las necesidades y las limitaciones particulares de cada contexto. Aunque los paquetes de *software* tengan sus puntos fuertes y débiles, siempre es preferible utilizarlos a llevar a cabo cálculos rutinarios a mano. Si la tecnología se implementa de una manera rigurosa, se obtienen investigaciones estadísticas más ricas y profundas que permitirán al alumnado apreciar plenamente la importancia de la estadística. En distintos sitios web se dispone de enlaces a herramientas de código abierto para aprender estadística y trabajar con ella (p. ej., CAUSE [https://www.causeweb.org/cause] y las de la ASA, s.f.-b); para consultar otros recursos, véase www.nctm.org/gaise.

El papel de la probabilidad en estadística

Tanto el profesorado como el alumnado tienen que comprender que la estadística y la probabilidad no son lo mismo. La estadística *utiliza* la probabilidad de la misma manera que la física utiliza el cálculo. El alumnado de estadística debe desarrollar una comprensión sólida de algunos de los conceptos fundamentales y de las aplicaciones de la probabilidad. Para calcular probabilidades, esta comprensión conceptual será más útil que las destrezas procedimentales. La mayoría de las veces, cuando se necesitan cálculos, las probabilidades se calculan a partir de distribuciones dadas. El alumnado puede usar la tecnología o las tablas para hacer estos cálculos.

[23] N. de T.: En el original se utiliza el término *computer science*, pero, en el ámbito hispanohablante, el equivalente *ciencia de la computación* se utiliza únicamente en niveles educativos superiores; en el nivel escolar, la denominación más habitual es *informática*.

Responder la pregunta de si podrían los resultados observados deberse únicamente al azar requiere entender la aleatoriedad del proceso que generó las observaciones. ¿Qué valores puede producir el azar?, ¿con qué frecuencia se dan esos valores? En Pfannkuch y Budgett (2017) se puede encontrar un ejemplo sobre cómo usar la tecnología para crear representaciones de la probabilidad. Diseñar y ejecutar simulaciones es una herramienta esencial para desarrollar esta comprensión.

Por ejemplo, ¿cómo serían los resultados de un examen de 20 preguntas de tipo verdadero/falso si un estudiante elige sus respuestas basándose únicamente en el azar? El alumnado debe darse cuenta de que existe una gran variabilidad en los resultados posibles. Cualquier resultado entre 0 y 20 respuestas correctas es posible. Sin embargo, no todos estos resultados son igualmente probables. El alumnado puede simular varios resultados posibles lanzando al aire una moneda no trucada 20 veces y utilizando el número de caras para representar el número de respuestas correctas del examen. De esta forma, verá que hay resultados que se dan con mayor frecuencia que otros. Todas las personas, salvo las más escépticas, llegarían a la conclusión de que alguien que haya respondido correctamente el 100 % de las preguntas de un examen bien diseñado conocía la materia. Por el contrario, alguien que haya respondido correctamente el 50 % de las respuestas tiene un desempeño similar al de otra que lo haya resuelto al azar.

Una concepción errónea habitual sobre la probabilidad es considerar que los sucesos con baja probabilidad nunca ocurren, y que aquellos que tienen una alta probabilidad suceden siempre. Por ejemplo, una candidata para la que se estima que tiene un 80 % de posibilidades de ganar unas elecciones no necesariamente ganará. En un universo hipotético en el que se llevan a cabo muchas elecciones y la candidata tiene un 80 % de posibilidades de ganar, perderá en una de cada cinco (20 %) de esas elecciones. De la misma manera, una predicción de lluvia del 5 % no significa que hoy no vaya a llover. Podrá llover o no; la predicción simplemente afirma que, en días como hoy, el 5 % de ellos llueve.

Consideremos la pregunta: ¿existe alguna relación entre la ingesta diaria de carne roja y sufrir un ataque al corazón? Predecir los resultados para cualquier individuo en particular tiene un alto grado de incertidumbre. En cambio, desde la estadística se responde a partir de las características de los grupos. La probabilidad aporta una manera de describir los riesgos y las predisposiciones para grupos numerosos de personas (no para los individuos de esos grupos).

En la inferencia estadística utilizamos la variabilidad debida al azar para cuantificar el grado de incertidumbre de una estimación. Por ejemplo, si, para estimar el porcentaje de habitantes de los Estados Unidos que usan lentes de contacto, el alumnado del nivel C se basa en una muestra de personas de la escuela, debe entender que el porcentaje correspondiente a su muestra podría ser muy diferente al porcentaje real de la población. Sin embargo, dado que su muestra no es aleatoria, no puede conocer el alcance de esa diferencia. Si, en cambio, se seleccionaran al azar a las personas de la población, entonces sí que se podría cuantificar la incertidumbre con un margen de error. Por ejemplo, el alumnado puede estimar en un 12 % el porcentaje de personas adultas de los Estados Unidos que usan lentes de contacto, con un margen de error de más o menos 5 puntos porcentuales. Una estudiante con una sólida comprensión conceptual de la probabilidad lo interpretará como que puede estar muy segura de que el porcentaje real de personas de la población que usa lentes de contacto está entre el 7 % y el 17 %. Pero, incluso así, no es imposible que el porcentaje real se encuentre fuera de este intervalo.

Dos conceptos importantes y relacionados entre sí derivados de la probabilidad son la probabilidad condicionada y la independencia. La probabilidad condicionada nos permite actualizar las probabilidades cuando se conoce más información. Por ejemplo, se puede concluir que existe un 60 % de posibilidades de que a un estudiante elegido al azar le guste la música rap. Pero quizás exista una asociación entre su nivel académico y sus preferencias musicales. En ese caso, se podría usar la probabilidad condicionada para explicar que, si elegimos al azar a un estudiante de 4.º curso de educación primaria, hay un 45 % de posibilidades de que prefiera el rap; pero si el estudiante elegido es de 6.º, entonces esa probabilidad es del 70 %. Dicho de otra manera, condicionar el conocimiento del curso del estudiante cambia la probabilidad de que a un estudiante elegido al azar le guste el rap. Cuando las probabilidades no se ven afectadas por el conocimiento condicionado, las variables son independientes. Si la preferencia musical es independiente del curso, entonces la probabilidad de encontrar a un amante del rap es la misma en 4.º que en 6.º de primaria. En este caso, la probabilidad de seleccionar aleatoriamente a un estudiante al que le guste la música rap, independientemente del curso en el que esté, sería del 60 %.

La noción de observaciones independientes es fundamental en estadística. En esencia, cuando las observaciones son independientes, nos proporcionan piezas únicas de información para nuestro conjunto de datos. Por ejemplo, supongamos que estamos pesando galletas para ponerlas a la venta y queremos confirmar que cada galleta pesa más o menos lo mismo. Si las galletas recién hechas se deshacen un poco en la balanza y dejan unos restos, el peso de las galletas que se pesen a continuación puede ser mayor que el de las anteriores. En términos de probabilidad, conocer el orden en el que se pesa una galleta afecta a la probabilidad de que esa galleta pese demasiado. En este caso, los pesos no son medidas independientes. Si, en cambio, la balanza se limpia y se recalibra después de cada uso, el valor registrado para una galleta es independiente del valor registrado para cualquier otra, de modo que se logra la independencia.

La práctica estadística moderna incluye la predicción en contextos multivariantes. En estos casos, el foco de la investigación es un individuo y no un grupo. Podría ser útil usar el historial dietético de una persona para estimar las posibilidades de que padezca un infarto en los próximos diez años. Basándonos en estas probabilidades, las personas o los objetos podrían clasificarse en grupos (p. ej., «sufrirá un ataque al corazón» o «no sufrirá un ataque al corazón») y los datos posteriores se utilizarán para evaluar la tasa de acierto (o de error) de esas clasificaciones.

Aspectos esenciales de cada componente

I. Formular preguntas de investigación estadística

- → Formular preguntas de investigación estadística multivariante y determinar de qué manera se pueden recoger y analizar los datos para dar una respuesta.
- → Plantear preguntas de investigación estadística de resumen, comparación y asociación para encuestas, estudios observacionales y experimentos que usen datos primarios o secundarios.
- → Plantear preguntas de investigación de inferencia estadística sobre la causalidad y la predicción.

II. Recoger/considerar los datos

- → Aplicar un plan adecuado de recogida de datos primarios o selección de datos secundarios para la pregunta de investigación estadística de interés.
- → Distinguir entre encuestas, estudios observacionales y experimentos.
- → Entender qué constituye una buena práctica para diseñar un estudio muestral, un experimento y un estudio observacional.
- → Entender el papel de la selección aleatoria en los estudios muestrales y el efecto del tamaño muestral en la variabilidad de las estimaciones.
- → Entender el papel de la asignación aleatoria en los experimentos y sus implicaciones en las interpretaciones causa-efecto.
- → Entender los problemas de los sesgos y de las variables de confusión en los estudios observacionales, y sus implicaciones para la interpretación.
- → Entender las prácticas de manejo de datos que mejoran la reproducibilidad y aseguran un uso ético, incluidas las descripciones de las alteraciones, y comprender cuándo los datos pueden contener información sensible.
- → Entender cómo la preocupación por la privacidad y por los seres humanos puede afectar a la recogida y la distribución de los datos.
- → Entender que, en algunas circunstancias, los datos recogidos o considerados podrían no ser generalizables a la población objetivo, o que estos datos pueden ser toda la población.

III. Analizar los datos

- → Usar la tecnología para dividir y filtrar conjuntos de datos y transformar variables, incluyendo el suavizado de series temporales.
- → Identificar formas adecuadas de resumir datos cuantitativos o cualitativos usando tablas, representaciones gráficas y resúmenes numéricos estadísticos, lo que incluye el uso de la desviación típica como medida de variabilidad, y el diagrama de caja modificado para identificar valores atípicos.
- → Resumir y describir relaciones entre múltiples variables.
- → Entender cómo se usan las distribuciones muestrales (desarrolladas mediante simulación) para describir la variabilidad entre muestras de los estadísticos muestrales.
- → Desarrollar simulaciones para determinar distribuciones muestrales aproximadas y calcular p-valores a partir de esas distribuciones.
- → Describir la asociación entre dos variables cualitativas, usando medidas como la diferencia de proporciones o el riesgo relativo.
- → Describir la relación entre dos variables cuantitativas mediante la interpretación del coeficiente de correlación de Pearson y de la recta de regresión de mínimos cuadrados.
- → Usar simulaciones para investigar las asociaciones entre dos variables cualitativas y para comparar grupos.
- → Construir intervalos de predicción e intervalos de confianza a fin de determinar valores posibles para una observación predicha o una característica de la población.

IV. Interpretar los resultados

- → Utilizar la evidencia estadística obtenida en los análisis para responder las preguntas de investigación estadística y comunicar los resultados por medio de informes y presentaciones más formales.
- → Evaluar e interpretar el impacto de los valores atípicos en los resultados.
- → Entender qué significa que un resultado o una estimación de una característica de la población sea posible o no en comparación con la variación debida al azar.
- → Interpretar el margen de error asociado a una estimación de una característica de la población.
- → Reconocer la existencia de valores perdidos y entender cómo esos valores pueden añadir sesgos al análisis.
- → Usar el pensamiento multivariante para entender cómo influyen unas variables en otras.
- → Comunicar a otras personas el razonamiento estadístico y sus resultados en una variedad de formatos (verbal, escrito, visual).
- → Entender cómo interpretar adecuadamente los p-valores simulados.

Ejemplo 1: Revisión del nivel B. Los pinzones de Darwin. De la desviación absoluta media a la desviación típica

Formular preguntas de investigación estadística

En el nivel B, el alumnado examinó el conjunto de datos secundarios con respecto a los pinzones al considerar la pregunta de investigación estadística:

Entre el pinzón cactero y el pinzón terrestre mediano de las islas Galápagos, ¿cuál de ellos suele tener el pico más largo?

El alumnado del nivel C puede revisar esta misma cuestión, pero ampliando el análisis con respecto al efectuado en el nivel B.

Recoger/considerar los datos

Se puede utilizar el mismo conjunto de datos que en el nivel B, basado en una muestra de 101 pinzones (véase la tabla 15), con 59 pinzones terrestres medianos (PTM/*fortis*) y 42 pinzones cacteros (PC/*scandens*).

Tabla 15. Tabla de resumen de las variables disponibles

101 pinzones en total	
Nombre de las variables	**Descripción**
Anilla	N.º del pinzón
Especie	PC o PTM
Longitud del pico	Longitud del pico (mm)
Profundidad del pico	Profundidad del pico (mm)
Año	Año en que se registró la observación
Sequía	Antes o después de la sequía de 1977

La anilla representa el número de identificación de cada pinzón. La especie es una variable cualitativa con dos categorías: pinzón cactero (PC) y pinzón terrestre mediano (PTM). La longitud del pico es una variable cuantitativa continua que representa la medida de la longitud desde la base del pico hasta la punta, medida en milímetros (mm). La profundidad del pico es una variable cuantitativa continua que representa la altura de la parte más amplia del pico, también medida en milímetros. La variable «Año» es una variable cuantitativa discreta que identifica el año en que se hicieron las observaciones. Por último, la variable «Sequía» es una variable cualitativa con dos categorías que indican si las observaciones fueron registradas antes o después de 1977, año en que hubo una gran sequía.

La recogida de datos se diseñó con la idea de tener 30 pinzones de cada especie, anteriores y posteriores a la sequía; sin embargo, la muestra de pinzones cacteros recogida con posterioridad a la sequía fue menor.

Analizar los datos

En el nivel B, el alumnado describió las distribuciones de la profundidad del pico de las muestras de pinzones cacteros (PC) y pinzones terrestres medianos (PTM) como aproximadamente simétricas y unimodales (véase la figura 50). Debido a que las distribuciones eran aproximadamente simétricas, sin valores atípicos extremos, se eligió la media como medida de tendencia central.

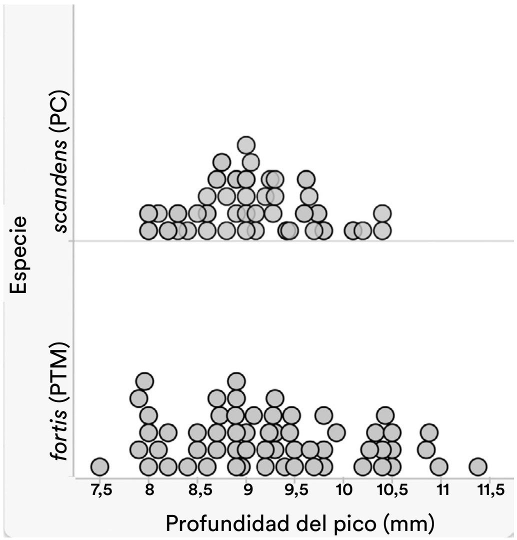


Figura 50. Profundidad del pico de las especies PC/*scandens* y PTM/*fortis*

Asimismo, el alumnado calculó la desviación absoluta media (DAM) como medida de variabilidad de la media de la distribución de una variable cuantitativa. La DAM es la distancia promedio entre cada dato y la media. Este cálculo usa el valor absoluto de (valor observado – media) para que todas las distancias sean positivas. Es decir:

$$\text{DAM} = \frac{\text{distancia total a la media para todos los valores}}{\text{números de datos}} = \frac{\sum|\text{valor observado} - \text{media}|}{n}$$

En el nivel C se descubre que otra forma de hacer que la diferencia (valor observado – media) sea positiva es elevándola al cuadrado. Si elevamos al cuadrado las diferencias, las sumamos, y las dividimos entre $n - 1$, el resultado se llama *varianza* (V),[24] que puede escribirse como:

$$V = \frac{\sum(\text{valor observado} - \text{media})^2}{n-1}$$

[24] N. de T.: En ocasiones, en el ámbito hispanohablante se distingue esta expresión (en la que se divide entre $n - 1$) de la que resulta cuando se divide entre n, y se denomina *cuasivarianza* a la primera y *varianza* a la segunda. En otros textos se las denomina, respectivamente, *varianza de la muestra* y *varianza de la población*. Consideramos que, en el ámbito de este documento, no es necesario un mayor detalle, ya que se explica la corrección en el párrafo siguiente, cuando se define la desviación típica (DT).

La varianza es el promedio de la distancia cuadrática de cada valor con respecto a la media, y resulta útil porque, cuando se consideran variables aleatorias independientes, sus varianzas se pueden sumar. No obstante, no se expresa en las mismas unidades que los datos dados. Al extraer la raíz cuadrada de la varianza, sin embargo, resulta una cantidad en las mismas unidades. Esta cantidad es la desviación típica (DT), que puede escribirse como:

$$DT = \sqrt{\frac{\sum(\text{valor observado} - \text{media})^2}{n-1}}$$

El alumnado puede darse cuenta de que el objetivo era obtener el promedio de la distancia cuadrática respecto a la media; entonces podría preguntarse por qué dividimos entre n – 1 y no entre n. Cuando se analiza una muestra de la población en lugar de toda la población, el denominador es n – 1 en lugar de n. Como la media muestral se usa para encontrar las diferencias en la muestra, entonces n – 1 diferencias determinan la última diferencia, ya que todas ellas deben sumar 0. Habitualmente, los valores observados en la muestra se sitúan, como promedio, más cerca de la media muestral que los valores de la población real. Por lo tanto, si se divide la suma entre n, la DT de la muestra subestimará la DT de la población. Para corregir esto matemáticamente, la suma se divide entre n – 1.

Tabla 16. Estadísticas de resumen de los pinzones

Especies	Media	DAM	DT
PC	9,07 mm	0,49 mm	0,62 mm
PTM	9,26 mm	0,73 mm	0,90 mm

El cálculo de la fórmula de la DT, con todos los elementos que intervienen, supone un reto mayor para el alumnado si lo comparamos con el de la DAM, y lo que se le suele escapar de este cálculo no intuitivo es cómo interpretarlo. La DAM puede actuar como puente o andamio para la DT, dado que es más intuitiva matemáticamente. Cuando el alumnado aprende a interpretar la DAM, desarrolla una comprensión conceptual que es similar a la que permite interpretar la DT.

Con respecto a los datos de los pinzones, la DAM de 0,73 mm (véase la tabla 16) puede interpretarse como hasta qué punto los valores de la profundidad del pico de los PTM se alejan de la media de 9,26 mm, como promedio. Dicho de otra manera, la profundidad del pico es generalmente 9,26 ± 0,73 mm (entre 8,53 mm y 9,99 mm). Puede aplicarse una interpretación similar para la DT: la profundidad del pico de los PTM difiere habitualmente de la media de 9,26 mm en ± 0,90 mm.

La DAM y la DT tendrán, en general, valores similares. En el caso de los pinzones PTM, se esperaría una DT mayor que la de los pinzones PC, dado que las observaciones de los PTM tienen un rango mayor y están menos concentradas en torno a la media que las observaciones de los PC. El alumnado del nivel C podrá apreciar, basándose en el cálculo, que la DT es más sensible a valores atípicos. La DAM es un estadístico más robusto, dado que es menos sensible a valores atípicos y puede ser más útil en algunas situaciones.

Entonces, ¿por qué la DT se usa más en la práctica que la DAM? Una de las razones principales es la relación matemática de la DT con las distribuciones en forma de campana, simétricas y unimodales. Una distribución especial de este tipo, la llamada *distribución normal*, tiene las siguientes características descritas por la regla empírica (véase la figura 51):

- Se espera que aproximadamente el 68 % de las observaciones se sitúen a menos de 1 DT de la media.
- Se espera que aproximadamente el 95 % de las observaciones se sitúen a menos de 2 DT de la media.
- Se espera que todas o casi todas las observaciones se sitúen a menos de 3 DT de la media.

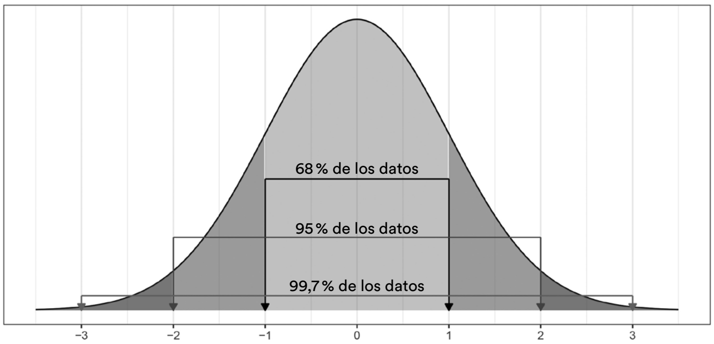


Figura 51. Distribución normal y regla empírica. (Esta figura se publicó originalmente, en inglés, en Hopfensperger [2020, fig. 14.3].)

Interpretar los resultados

Al igual que en el nivel B, el alumnado del nivel C llegaría a la conclusión de que no existe una diferencia significativa entre la profundidad del pico de los pinzones PTM y PC. Las distribuciones tienen centros y variabilidad similares, medidos por la media y las desviaciones típicas. Las distribuciones tienen formas parecidas y, en su mayoría, se solapan con muy poca separación.

Ejemplo 2: Revisión de los niveles A y B. La elección de la música para la fiesta escolar (continuación). Generalizar los resultados

Formular preguntas de investigación estadística

En el nivel A se introdujo un sondeo sobre las preferencias musicales, cuyo análisis consistió en el recuento de las respuestas de los estudiantes y la representación de los datos en un diagrama de barras. En el nivel B, el análisis se amplió para considerar las frecuencias relativas de las preferencias y cruzar las respuestas sobre dos estilos musicales, para exponerlas después en una tabla de doble entrada.

En ejemplos anteriores, se extraían los datos de una clase (o un curso), y las generalizaciones no se extendían formalmente más allá de esa clase (o curso). En el nivel C, el alumnado se plantea de qué manera puede generalizar a todo el centro las observaciones obtenidas de una muestra de unos pocos estudiantes de un centro educativo.

Un estudiante cree que, para la fiesta escolar, toda la música debería ser rap porque «a la mayoría del alumnado le gusta el rap». Para probar esta afirmación, plantea la siguiente pregunta de investigación estadística:

¿A la mayoría del alumnado le gusta el rap?

Dicho de otra manera:

¿A más del 50 % del alumnado de nuestro centro le gusta el rap?

Esta pregunta también podría plantearse como una afirmación:

A más del 50 % del alumnado de nuestro centro le gusta el rap.

Recoger/considerar los datos

Recordemos que el cuestionario del nivel B incluía las siguientes preguntas:

P1. Marca «Sí» para todos los tipos de música que te gusten entre los siguientes. Marca «No» para cualquiera que no te guste.

	Sí	**No**
Rap	☐	☐
Rock	☐	☐
Country	☐	☐
R&B	☐	☐
Pop	☐	☐
Clásica	☐	☐
Alternativa	☐	☐

P2. ¿Cuál es tu tipo de música favorito?

☐ Rap
☐ *Rock*
☐ *Country*
☐ *R&B*
☐ Pop
☐ Clásica
☐ Alternativa
☐ Otro

Para generalizar a todo el alumnado del centro, se necesita una muestra representativa de los estudiantes del centro. En este nivel C se seleccionó una muestra aleatoria simple de 60 estudiantes del centro a los que administrar el cuestionario. Los resultados se pueden generalizar a ese centro (pero no más allá), y las consideraciones en este nivel pueden centrarse en los principios básicos de la generalización, es decir, la inferencia estadística.

Analizar los datos

La pregunta de investigación estadística solo tiene en cuenta la variable «le gusta el rap». Dado que al 58 % del alumnado de la muestra le gustaba la música rap (lo que es más del 50 %), hay evidencia que respalda la afirmación de que a más del 50 % del alumnado de todo el centro le gusta el rap. Sin embargo, es posible que ese valor observado del 58 % se deba al azar a causa del muestreo aleatorio, y que el verdadero porcentaje de alumnado al que le gusta el rap sea menor o igual que el 50 % (en cuyo caso, la afirmación del estudiante no sería correcta).

Se podría usar una simulación para examinar hasta qué punto es posible (o probable) una proporción observada del 58 % de éxito (le gusta el rap) cuando la población tiene solo un 50 % de éxito. Podemos asumir una población hipotética con un 50 % de éxito, lo que se puede modelar lanzando monedas. El alumnado puede realizar esta simulación lanzando una moneda 60 veces. En grupo, pueden combinar sus resultados para formar una distribución de proporciones muestrales simuladas para investigar en cuántos casos se obtuvo un 58 % o más de caras. Deberían debatir si esto les conduce a creer que un resultado del 58 % de éxitos en una muestra es un resultado raro o un resultado habitual cuando el porcentaje real es del 50 %. A menos que el grupo sea muy muy numeroso, no tendrá muchas observaciones simuladas para extraer una conclusión fundamentada. Un *applet* o un *software* estadístico permitirá al alumnado producir muchas más muestras. Lanzando una moneda virtual 60 veces, anotando el número de aciertos y repitiéndolo muchas veces, se puede generar una distribución muestral aproximada, similar a la de la figura 52.

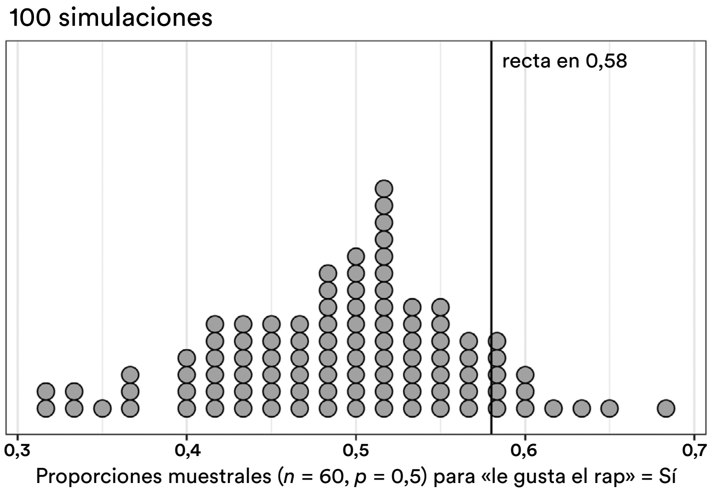


Figura 52. Distribución de las proporciones muestrales simuladas cuando la proporción de la población = 0,50

Interpretar los resultados

Según esta distribución de las proporciones muestrales simuladas, se dio una proporción muestral mayor o igual a 0,58 en 17 de 100 ocasiones (contando el número de puntos que se sitúan en 0,58 o a su derecha en el gráfico de puntos) simplemente por la variación debida al azar cuando la verdadera proporción de la población es de 0,50. Esto sugiere que un resultado de 0,58 no es un hecho muy inusual cuando se toman muestras de una población con el 0,50 como la verdadera proporción de alumnado al que le gusta la música rap. Por tanto, las evidencias para afirmar que a más del 50 % del alumnado le gusta el rap no son muy sólidas. Asumiendo que el valor del 50 % fuera correcto, la proporción de veces en las que el resultado observado alcanza o supera ese valor en las distintas repeticiones del experimento (0,17 en esta investigación) es una aproximación de lo que se llama *p-valor* (usar más de 100 repeticiones de la simulación proporcionaría una aproximación mejor). Un p-valor bajo habría indicado que, si la proporción de la población fuese 0,50, sería muy improbable que se hubiese observado una proporción muestral de 0,58 (indicada por la recta vertical en el gráfico).

En cambio, otra estudiante podría plantear la siguiente pregunta de investigación estadística:

¿A más del 40 % del alumnado de nuestro instituto le gusta el rap?

Esto llevaría a la afirmación:

A más del 40 % del alumnado de nuestro instituto le gusta el rap.

Para investigar la pregunta de esta estudiante, se deben seleccionar de manera repetida muestras de tamaño 60 a partir de una población hipotética con un 40 % de éxito. Así, la población observada de 0,58 puede compararse con las proporciones simuladas extraídas de una población cuya proporción es del 40 %. La figura 53 muestra los resultados de 100 muestras simuladas. El resultado observado incluye una muestra que produjo una proporción superior a 0,58. De esta manera, el p-valor es aproximadamente 0,01, que es muy bajo, lo que indica que no es probable que una población en la que al 40 % del alumnado le gusta la música rap hubiera producido una proporción muestral de 0,58 o mayor en un muestreo aleatorio de tamaño 60. Este p-valor bajo aporta una evidencia muy sólida para respaldar la afirmación, formulada por la estudiante, de que a más del 40 % del alumnado de todo el instituto le gusta la música rap.

Otra forma de explicar lo anterior es decir que 0,5 es un posible valor verosímil para la proporción poblacional real (p) basándonos en la evidencia de la muestra, mientras que 0,4 no lo es. Se puede utilizar un intervalo de confianza para describir un conjunto de valores posibles, mediante un margen de error. El margen de error para un nivel de confianza del 95 % en una proporción muestral es:

$$2 \cdot \sqrt{\frac{p(1-p)}{n}}$$

La fórmula puede estar motivada por la exposición repetida del alumnado a simulaciones de las que extrae muchas muestras aleatorias a partir de una población con un valor de p conocido y examina la distribución empírica resultante de las proporciones muestrales. Una investigación más profunda mostrará al alumnado que el 95 % de las proporciones muestrales en la distribución simulada que se ve en la figura 53 (donde la proporción de la población es de 0,4) se sitúa dentro de una distancia de

$$2 \cdot \sqrt{\frac{(0,4)(0,6)}{60}} \approx 0,126$$

del valor real de p.

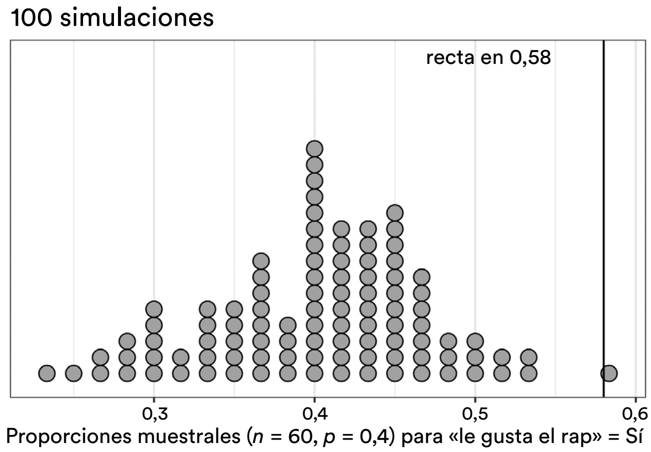


Figura 53. Distribución de las proporciones muestrales simuladas cuando la proporción de la población = 0,40

Sin embargo, en la mayoría de los casos, no se conoce el valor de p y, por tanto, tenemos que estimarlo con la proporción muestral $\hat{p}$.

$$2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

En nuestro estudio sobre las preferencias musicales, se desconoce la verdadera proporción del alumnado al que le gusta el rap. Nuestra proporción muestral ($\hat{p}$ = 0,58) es nuestra «mejor estimación» del que podría ser su valor, por lo que se puede estimar que el margen de error es:

$$2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2 \cdot \sqrt{\frac{(0,58)(0,42)}{60}} \approx 0,127$$

Así, cualquier proporción entre 0,58 – 0,127 = 0,453 y 0,58 + 0,127 = 0,707 puede considerarse un valor posible de la verdadera proporción de alumnado del centro al que le gusta el rap. Nótese que 0,5 está dentro del intervalo, mientras que 0,4 no. Esto resulta congruente con nuestra conclusión previa: 0,5 es un valor posible para la verdadera proporción de la población, pero 0,4 no lo es.

Los intervalos de confianza y los márgenes de error se suelen proporcionar, en los medios de comunicación, cuando se informa de resultados de votaciones u otros estudios. En el contexto de la educación, los resultados de los exámenes normalmente también se muestran con intervalos y márgenes de error de tipo ±. Es importante que el alumnado del nivel C entienda cómo interpretar intervalos de confianza, porque es parte de su alfabetización estadística.

Ejemplo 3: La elección de la música para la fiesta escolar (continuación). Inferencia sobre la asociación

Formular preguntas de investigación estadística

Otro tipo de pregunta de investigación que podría hacerse en relación con las preferencias musicales del alumnado es:

A quienes les gusta la música rock*, ¿les suele gustar más la música rap que a quienes no les gusta el* rock*?*

Dicho de otra manera: *¿existe una asociación entre el interés por la música* rock *y por el rap?*

Recoger/considerar los datos

Para responder esta pregunta, se pueden usar los mismos datos de la muestra aleatoria de 60 estudiantes. El alumnado debe identificar que las variables relevantes de esta recogida de datos son «le gusta el rap» y «le gusta el *rock*».

Analizar los datos

Para el alumnado de esta clase, la asociación entre el gusto por el *rock* y el gusto por el rap puede resumirse en una tabla de doble entrada, como la tabla 17.

Tabla 17. Tabla de frecuencias de doble entrada

		Rock		
		Sí	**No**	**Total**
Rap	Sí	30	5	35
	No	7	18	25
	Total	37	23	60

Según la tabla 17, a un total de 37 estudiantes encuestados les gusta la música *rock*. Dentro de este grupo, la proporción de estudiantes a quienes también les gusta la música rap es de 30/37 = 0,81. De entre los 23 estudiantes a quienes no les gusta el *rock*, 5/23 = 0,22 es la proporción de a quienes les gusta el rap. La diferencia entre estas dos proporciones muestrales (0,59) sugiere que puede haber una fuerte asociación entre el gusto por el *rock* y el gusto por el rap. Pero ¿podría esa diferencia de porcentaje relativamente grande observada en la muestra deberse simplemente al azar (es decir, ser solo una consecuencia del muestreo aleatorio)?

Para dar respuesta a esta pregunta, el alumnado del nivel C debe tener en cuenta una población hipotética en la que no haya asociación entre las dos variables. En esa población, la proporción de alumnado al que le gusta el rap sería la misma entre el alumnado al que le gusta el *rock* y entre el que no le gusta el *rock*. Por tanto, se espera que la proporción de alumnado al que le gusta el rap, de entre los 37 estudiantes a quienes les gusta el *rock*, será cercana a la proporción de alumnado al que les gusta el rap, de entre los 23 estudiantes a quienes no les gusta el *rock*. Básicamente, si no hay asociación en la población, se espera que la diferencia entre estas dos proporciones muestrales sea aproximadamente 0.

Para simular esta situación, el alumnado puede utilizar 60 cartas que representen cada una de las observaciones de la muestra. Puede marcar 35 de esas cartas como «le gusta el rap» y las 25 restantes como «no le gusta el rap». El alumnado aleatoriza sus cartas barajándolas bien y repartiendo 37 en un montón y el resto en otro. El primer montón representa a quienes les gusta el *rock;* y el segundo, a quienes no les gusta el *rock*.

Esta ronda de barajar y repartir las cartas simula una situación en la que no existe asociación entre el gusto por el rap y el gusto por el *rock*. Que una carta marcada como «le gusta el rap» termine en el montón de «le gusta el *rock*» o en el de «no le gusta el *rock*» está completamente determinado por el azar. La diferencia entre las proporciones de las cartas de «le gusta el rap» en ambos montones simula la diferencia de proporciones en una población en la que no existe asociación entre las dos variables. Al repetir este proceso de barajar y repartir las cartas muchas veces y anotar la diferencia de proporciones de cartas «le gusta el rap» en los dos montones, el alumnado puede construir una distribución muestral aproximada.

El proceso de barajar y repartir las cartas es lento y engorroso, pero importante como forma inicial de simular, y permite que el alumnado experimente y visualice el papel de la aleatoriedad en «la variación debida al azar». Una vez que se haya familiarizado con el proceso de simulación, se puede utilizar la tecnología para repetir el proceso 100 veces de una manera rápida. La distribución muestral aproximada resultante tendrá un aspecto similar al que puede verse en la figura 54.

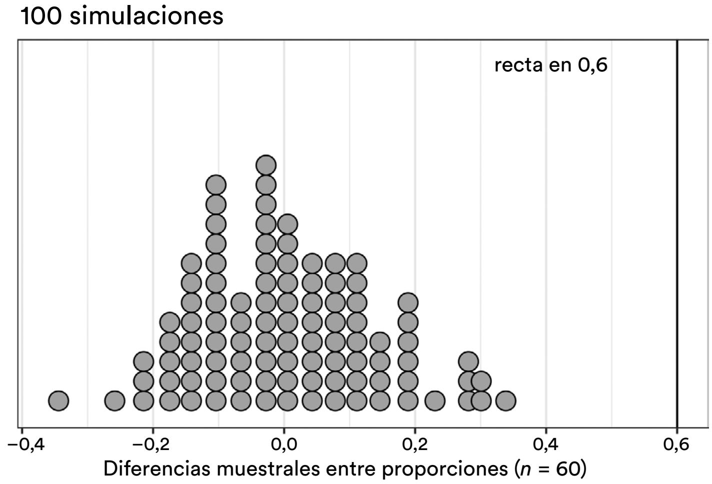


Figura 54. Gráfico de puntos con la distribución de la diferencia entre dos proporciones en la muestra simulada

Interpretar los resultados

La diferencia observada en las proporciones a partir de los datos de la muestra, 0,59, nunca se alcanzó cuando se hicieron 100 repeticiones. Así pues, el p-valor estimado es 0/100 = 0, lo que indica claramente que la diferencia observada no puede atribuirse de manera razonada únicamente al azar. Por tanto, existe evidencia sólida de una asociación real entre el gusto por la música *rock* y el gusto por la música rap entre el alumnado del centro.

Ejemplo 4: El efecto de la luz en el crecimiento de plántulas de rábano. Experimentos

Formular preguntas de investigación estadística

Un grupo de estudiantes de Biología planteó la siguiente pregunta de investigación estadística:

¿Cuál es el efecto de los diferentes tiempos de exposición a la luz y a la oscuridad en el crecimiento de las plántulas de rábano?

A continuación, el alumnado diseña y lleva a cabo un experimento para investigar la cuestión.

Recoger/considerar los datos

No se pueden investigar todas las posibles exposiciones a la luz y a la oscuridad en un solo experimento, por lo que el alumnado del nivel C podría decidir centrar el experimento en tres tratamientos: 24 horas de luz (*luz*), 12 horas de luz y 12 horas de oscuridad (*combinado*) y 24 horas de oscuridad (*oscuridad*). Así se cubren los dos casos extremos y uno intermedio.

Con la ayuda de la profesora, la clase decide usar bolsas de plástico como cámaras de crecimiento. Estas bolsas permiten observar y medir la germinación de las semillas sin obstaculizarlas. Se introdujeron dos capas de servilletas de papel húmedas en una bolsa de plástico transparente, con una línea de grapas a aproximadamente 1/3 de la parte inferior de la bolsa (véase la figura 55) para mantener la servilleta en su sitio y proporcionar una costura en la que sujetar las semillas de rábano.

Para el estudio se disponía de 120 semillas. El tamaño de las cámaras de crecimiento permitía introducir solo 30 semillas. La clase decidió dar uso a las semillas extra y crear cuatro cámaras de crecimiento: una para el tratamiento con *luz*, otra para el tratamiento *combinado* y dos para el tratamiento con *oscuridad*. Se tomaron, al azar, 30 semillas y se colocaron a lo largo de la línea grapada de la bolsa del tratamiento con *luz*. A continuación, se tomaron, también al azar, otras 30 semillas, de las 90 restantes, y se colocaron en la bolsa del tratamiento *combinado*. Finalmente, se escogieron al azar 30 semillas, de las 60 restantes, y se introdujeron en una de las bolsas del tratamiento con *oscuridad*. Las últimas 30 se colocaron en la otra bolsa del tratamiento con *oscuridad*. El alumnado procuró que las cámaras de crecimiento fueran lo más parecidas posible. Contar con dos cámaras de crecimiento bajo las mismas condiciones, permitió comparar los resultados y asegurarse de que el manejo de las semillas fuera similar. Después de tres días se midieron y se registraron las longitudes (en milímetros) de las plántulas de rábano procedentes de las semillas en germinación.

Figura 55. Bolsa de crecimiento del experimento de las semillas para 30 plántulas

Inicialmente, los datos se registraron en un formato resumido similar al que se muestra en la tabla 18, con los valores ordenados. En todos los tipos de tratamiento hubo semillas que no germinaron y que se registraron con una «x» y se consideraron valores perdidos. Por tanto, hay un total de 114 observaciones (28 para *luz*, 28 para *combinado* y 58 para *oscuridad*). En el nivel C se debe animar al alumnado a debatir si la omisión de las semillas que no germinaron podría añadir sesgo a las conclusiones. Aquí se decidió que, dado que el número de semillas que no habían logrado germinar era más o menos el mismo en cada categoría, los valores perdidos probablemente se debieran «al azar», por lo que no afectarían a las conclusiones. Alternativamente, si todos los valores perdidos estuvieran en una categoría, esto sugeriría que las condiciones de dicha categoría obstaculizaron el crecimiento, por lo que esos datos perdidos no se deberían «al azar».

Tabla 18. Longitudes de las plántulas de rábano tras tres días (ordenadas)

Tipo de tratamiento	Longitudes de las plántulas (mm)
1-*luz*	x, x, 2, 3, 5, 5, 5, 5, 5, 7, 7, 7, 8, 8, 8, 9, 10, 10, 10, 10, 10, 10, 10, 10, 14, 15, 15, 20, 21, 21
2-*combinado*	x, x, 3, 4, 5, 9, 10, 10, 10, 10, 10, 11, 13, 15, 15, 15, 17, 20, 20, 20, 20, 20, 21, 21, 22, 22, 23, 25, 25, 27
3-*oscuridad*	x, 5, 8, 8, 10, 10, 14, 15, 15, 15, 20, 20, 20, 20, 20, 22, 25, 25, 25, 25, 26, 30, 30, 35, 35, 35, 35, 36, 37, 38
3-*oscuridad*	x, 5, 8, 8, 10, 10, 10, 11, 14, 15, 15, 15, 16, 20, 20, 20, 20, 20, 24, 25, 29, 30, 30, 30, 30, 31, 33, 35, 35, 40

La estructura de la tabla 18 resultó más conveniente para registrar las observaciones. No obstante, para los objetivos de análisis, los datos también se podrían representar en un formato extendido con cada observación (semilla) en una fila independiente y cada variable en una columna. La tabla 19 muestra las tres primeras y las tres últimas filas de los datos en formato extendido.

Tabla 19. Listado en formato extendido de las longitudes de las plántulas de rábano

Semilla #	Bolsa de crecimiento	Tratamiento	Longitud (mm)
1	1	1-*luz*	x
2	1	1-*luz*	x
3	1	1-*luz*	2
…	…	…	…
118	4	3-*oscuridad*	35
119	3	3-*oscuridad*	35
120	4	3-*oscuridad*	

Las unidades de observación son semillas individuales, la bolsa de crecimiento es una variable cualitativa que indica en qué bolsa está la semilla (1, 2, 3 o 4), el tratamiento es otra variable cualitativa que indica qué tratamiento están recibiendo las semillas (1-*luz*, 2-*combinado* o 3-*oscuridad*), y la longitud es una variable cuantitativa que mide la longitud en milímetros de las plántulas procedentes de las semillas.

La pregunta de investigación estadística planteada es de causalidad; la hipótesis implícita es que, si existen diferencias, están causadas por los niveles de luz. El diseño experimental, dado que emplea una asignación aleatoria a los niveles de tratamiento, permite concluir que las diferencias, si no se deben al azar, están causadas por los niveles de luz. Pero el primer paso es descartar la posibilidad de que las diferencias se deban únicamente al azar.

Analizar los datos

Como se expuso en los niveles A y B, es importante examinar en detalle los datos; y un buen primer paso para el análisis de datos cuantitativos como estos es representarlos gráficamente. Los diagramas de caja son ideales para comparar los datos para la variable de respuesta «Longitud del rábano» obtenidos para más de un tratamiento, como se ve en la figura 56. Tanto el centro como la variabilidad de las distribuciones aumentan a medida que se incrementa el tiempo de oscuridad. Hay tres potenciales valores atípicos (uno en 20 mm y dos en 21 mm) en los datos del tratamiento 1-*luz*.

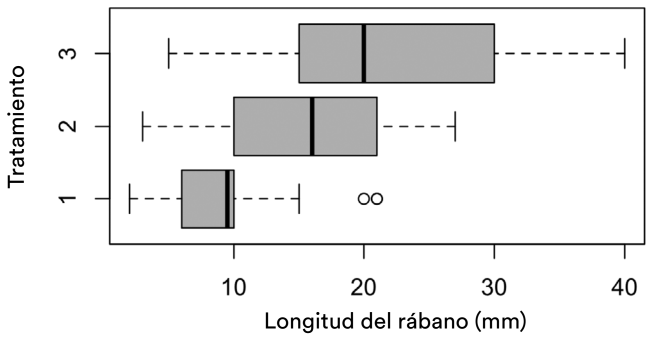


Figura 56. Diagrama de caja que muestra el crecimiento bajo diferentes condiciones: 1-*luz*, 2-*combinado* y 3-*oscuridad*

Las estadísticas de resumen para estos datos se recogen en la tabla 20.

Los experimentos se diseñan para comparar los efectos de los tratamientos. La pregunta original sobre el efecto de los diferentes periodos de luz y oscuridad en el crecimiento de las plántulas de rábano podría convertirse en preguntas de análisis comparativo entre las medias de cada tratamiento. Dos de esas preguntas podrían ser:

Tabla 20. Estadísticas de resumen según tratamiento

Tratamiento	n	Media	Mediana	Desv. típ.
1-*luz*	28	9,64	9,50	5,03
2-*combinado*	28	15,82	16,00	6,76
3-*oscuridad*	58	21,86	20,00	9,75

¿Hay evidencia de que el grupo con 12 horas de luz y 12 horas de oscuridad (tratamiento 2) tiene una longitud media significativamente mayor que el grupo con 24 horas de luz (tratamiento 1)?

¿Hay evidencia de que el grupo con 24 horas de oscuridad (tratamiento 3) tiene una longitud media significativamente mayor que el grupo con 12 horas de luz y 12 horas de oscuridad (tratamiento 2)?

A partir de los diagramas de caja y las estadísticas de resumen, parece que las medias de los grupos difieren. Esto da pie a la siguiente pregunta de análisis:

¿Son las diferencias observadas entre las medias lo bastante grandes como para descartar que se deban al azar?

La longitud media de las semillas en el grupo con el tratamiento 2-*combinado* es 6,2 mm mayor que la del grupo con el tratamiento 1-*luz*. Aunque hay una diferencia de 6,2 mm, puede no ser lo bastante grande como para descartar que se deba al azar, así que tal vez no podamos afirmar que exista un efecto del tratamiento. Esta diferencia observada puede deberse, simplemente, a que una de las bolsas haya tenido la buena suerte de recibir un número elevado de semillas buenas. También puede haberla causado una distribución desigual de agua o de luz entre los grupos de tratamiento. Pero si es probable que una diferencia tan grande (6,2 mm) se deba solo a la aleatorización de las semillas en los tratamientos, deberíamos ver diferencias similares con bastante frecuencia si volviéramos a repetir la aleatorización de las medidas y calculásemos, cada vez, la nueva diferencia entre las medias observadas. El alumnado puede simular esta nueva aleatorización escribiendo las longitudes de las 56 semillas en tarjetas, barajándolas y repartiéndolas en dos montones de 28 tarjetas cada uno. El primer montón representa el grupo con tratamiento 1-*luz*; y el segundo, el grupo con tratamiento 2-*combinado*. A continuación, puede calcular la diferencia entre las medias de ambos montones. Esta diferencia simulada se debe enteramente al azar. Repitiendo este proceso muchas veces, el alumnado puede comprender cómo varía la diferencia entre las medias cuando el tratamiento no tiene ningún efecto en el crecimiento de las plántulas.

Hacer esta simulación, incluso una sola vez, es tedioso y lento, pero será útil para el alumnado cuando realice la simulación mediante la tecnología. La figura 57 se elaboró utilizando herramientas tecnológicas para mezclar las medidas de crecimiento para los tratamientos 1-*luz* y 2- *combinado*, y separarlas aleatoriamente en dos grupos de 28. Se registró la diferencia entre las medias y se repitió el proceso 200 veces.

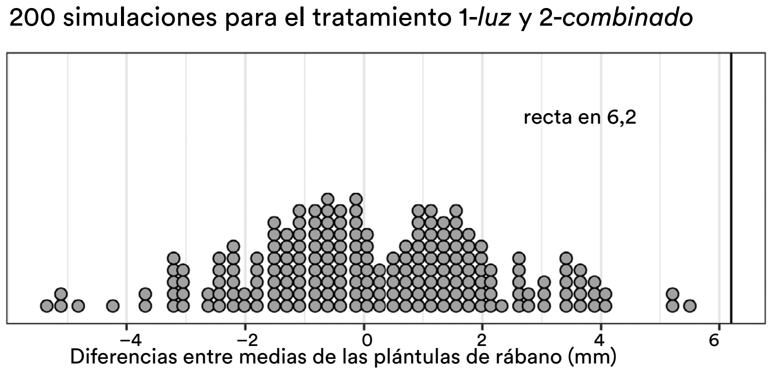


Figura 57. Distribución de la nueva aleatorización de la diferencia entre las medias para el tratamiento 1-*luz* y el tratamiento 2-*combinado*

En esta simulación de 200 nuevas aleatorizaciones nunca se sobrepasó la diferencia observada de 6,2 mm con un p-valor aproximado de 0/200. Esto proporciona evidencia sólida contra la hipótesis de que la diferencia entre las medias de los tratamientos 1 y 2 se debe únicamente al azar.

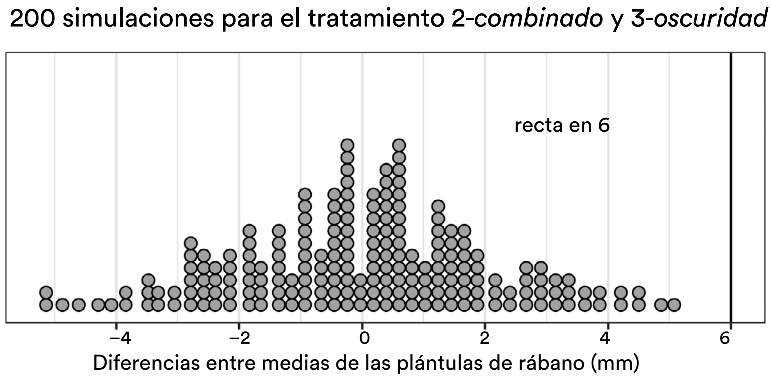


Figura 58. Distribución de la nueva aleatorización de la diferencia entre las medias para el tratamiento 2-*combinado* y el tratamiento 3-*oscuridad*

Se siguió el mismo procedimiento en una comparación de las medias de los tratamientos 2 y 3: mezclar las medidas de crecimiento de los tratamientos 2-*combinado* y 3-*oscuridad*, separarlos aleatoriamente en dos grupos de 28 y 58 medidas, registrar la diferencia entre las medias de ambos grupos y repetir el proceso 200 veces. La diferencia observada de 6 mm no se superó en ninguno de los 200 ensayos (véase la figura 58), lo que proporciona evidencia sólida de que la diferencia observada entre las medias de los tratamientos 2 y 3 no se debe únicamente al azar.

Interpretar los resultados

En resumen, dos comparaciones por pares entre los tres grupos de tratamiento muestran diferencias en el crecimiento medio que no pueden explicarse razonablemente por la asignación aleatoria de las semillas a las bolsas. Esto proporciona pruebas convincentes de un efecto del tratamiento: a más horas de oscuridad, mayor crecimiento de la plántula, al menos para estos tres periodos de luz frente a oscuridad. La asignación aleatoria permite que el alumnado concluya que las diferencias de crecimiento fueron causadas por los distintos niveles de luz.

Este análisis es un ejemplo de comparaciones múltiples. Para concluir que más horas de oscuridad producen un mayor crecimiento, se plantearon y se respondieron dos preguntas de investigación estadística. Sin embargo, el alumnado del nivel C debe tener en cuenta que los análisis realizados podrían ser problemáticos. Solo se llevaron a cabo dos comparaciones por pares (se compararon entre sí los tratamientos 1 y 2 y los tratamientos 2 y 3), pero había tres pares de condiciones de tratamiento diferentes. Para responder completamente las preguntas de investigación, se debe animar al alumnado a explorar la pareja comparativa que falta. Quienes superen el nivel C también pueden analizar los posibles inconvenientes de las comparaciones múltiples por pares. Existen métodos más avanzados, pero quedan fuera del ámbito del nivel C; se pueden encontrar en cursos de nivel universitario.

Hay que animar al alumnado a que profundice en la interpretación, relacionándola con lo que ya se sabe sobre el fenómeno o el problema estudiado. *¿Por qué las plántulas crecen más rápido en la oscuridad?* El profesorado de Biología podría discutir cómo se han adaptado las plantas para germinar hacia la luz (por encima del suelo) lo más rápido posible. La plántula no puede hacer la fotosíntesis en la oscuridad y utiliza la energía almacenada en la semilla para potenciar el crecimiento. Cuando está expuesta a la luz, deja de dedicar su energía a crecer en longitud y pasa a dedicarla a producir clorofila y aumentar el tamaño de sus hojas. Estos cambios permiten que la planta sea autosuficiente y empiece a producir su propio alimento. A pesar de que el crecimiento en longitud del tallo se ralentiza, el crecimiento en diámetro del tallo y el crecimiento del tamaño de las hojas aumentan. Las plántulas que continúan creciendo en la oscuridad son escuálidas y amarillas, con hojas pequeñas y amarillas. Las plántulas cultivadas expuestas a la luz son de un color verde vibrante, con hojas grandes y gruesas, y tallos cortos.

Ejemplo 5: La consideración de las medidas en el diseño de prendas de vestir. Regresión lineal

Formular preguntas de investigación estadística

Al diseñar ropa, se tienen que crear los patrones de manera que las prendas puedan ser adecuadas para los compradores. Para que la ropa se adapte adecuadamente a las personas, quien realiza el diseño debe comprender la relación entre las dimensiones de varias partes del cuerpo. Por ejemplo, una diseñadora de camisas debe tener en cuenta la longitud del brazo con respecto a la del torso. Supongamos que una clase de Estadística recibe la tarea de crear un patrón para el vestuario del coro de la obra de teatro del instituto. El vestuario consiste en túnicas que caen hasta el suelo, por lo que los pies no se verían. Las túnicas están diseñadas de forma que la manga del antebrazo se ajuste firmemente al brazo, desde el codo hasta la mano. El alumnado plantea la siguiente pregunta de investigación estadística:

¿Qué relación hay entre la longitud del antebrazo de una persona y su estatura? ¿Se puede usar la longitud del antebrazo para predecir la estatura?

Recoger/considerar los datos

El alumnado de la clase de Estadística decide usar sus propias medidas corporales para hacer el patrón de las túnicas. Se toman las medidas en casa para evitar posibles situaciones incómodas en el aula. En el nivel C debe reconocer que medir solo a los estudiantes de su clase puede no ser generalizable a todo el centro; no obstante, sí que proporciona una base para el corte del patrón. Una consideración importante aquí es acordar la definición de «antebrazo» antes de empezar a tomar medidas. Se determina que el antebrazo de la túnica se mide desde el pliegue del codo por la parte interna del brazo hasta el final de la muñeca. Los datos (en centímetros) obtenidos por el alumnado se muestran en la tabla 21. En cada fila, cada par de columnas representa a un estudiante.

Tabla 21. Estaturas frente a longitudes del antebrazo

Antebrazo (cm)	Estatura (cm)	Antebrazo (cm)	Estatura (cm)
45,0	180,0	41,0	163,0
44,5	173,2	39,5	155,0
39,5	155,0	43,5	166,0
43,9	168,0	41,0	158,0
47,0	170,0	42,0	165,0
49,1	185,2	45,5	167,0
48,0	181,1	46,0	162,0
47,9	181,9	42,0	161,0
40,6	156,8	46,0	181,0
45,5	171,0	45,6	156,0
46,5	175,5	43,9	172,0
43,0	158,5	44,1	167,0

Si el alumnado no está familiarizado con las unidades del sistema métrico decimal,[25] puede querer examinar los datos en unidades que le resulten más conocidas. Para ello, se transforman los datos. Mediante el análisis dimensional pueden crearse nuevas variables que transformen las unidades de centímetros a pulgadas, multiplicando los valores «antiguos» por 0,393701 pulgadas/cm (o dividiendo entre 2,54 cm/pulgada).

Analizar los datos

Un buen primer paso en cualquier análisis es representar gráficamente los datos (véase la figura 59). El diagrama de dispersión indica que sería razonable modelar la relación entre la estatura y la longitud del brazo mediante una tendencia lineal. En el nivel B se consideraba la razón de recuento de cuadrantes (RRC) como un resumen numérico para cuantificar la fuerza de la relación entre dos variables cuantitativas. La RRC se basa en el recuento de puntos en cada uno de los cuatro cuadrantes del diagrama de dispersión definidos por las rectas de las medias de cada una de las dos variables. En el nivel C, el análisis pasa a utilizar el coeficiente de correlación de Pearson (r), que tiene en cuenta la distancia de los puntos a las rectas de las medias. Estas son una recta vertical y otra horizontal situadas en la media de las medidas de los antebrazos y la media de las medidas de las estaturas, respectivamente.

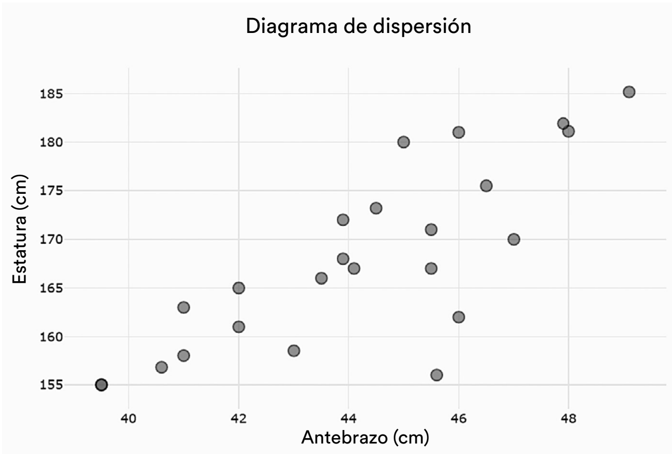


Figura 59. Diagrama de dispersión sin la recta de regresión

De manera similar a la RRC, la r de Pearson es adimensional y siempre está entre −1 y +1, ambos inclusive. Kader y Franklin (2008) desarrollan este aspecto con más detalle. La r de Pearson obtenida con la tecnología para estos datos es de 0,8. Esto indica que hay una relación lineal positiva y fuerte entre la estatura y la longitud del antebrazo.

Si la «nube» de puntos de un diagrama de dispersión tiene una forma lineal (un óvalo alargado o una forma similar a una salchicha en vez de un patrón más circular), una recta podría ser un modelo realista de la relación entre las variables en estudio. El alumnado del nivel C debe familiarizarse con la recta de regresión de mínimos cuadrados (mencionada a lo largo de este ejemplo como «recta de mínimos cuadrados»), que es la recta que minimiza la suma de los residuos al cuadrado. Los residuos se definen como las diferencias en vertical (eje Y) entre los puntos del diagrama de dispersión y la recta de mínimos cuadrados. Se pueden utilizar *applets* para mostrar el concepto de recta de mínimos cuadrados y los residuos. La recta de mínimos cuadrados también pasa por el punto donde se cruzan las rectas de las medias (media de x, media de y) de la nube de puntos.

La ecuación de la recta de mínimos cuadrados (calculada con *software* estadístico) de esta relación que se muestra en el diagrama de dispersión de la figura 60 es:

$$\text{altura predicha} = 45{,}8 + 2{,}76\ (\text{longitud del antebrazo})$$

El gráfico que se presenta debajo del diagrama de dispersión muestra los residuos (véase la figura 60).

Los residuos indican una variación en torno a la recta de mínimos cuadrados, medida a través de su desviación típica.

[25] N. de T.: Como se señaló con anterioridad, en los Estados Unidos no es común el uso del sistema métrico decimal. El alumnado de países donde es habitual el sistema métrico decimal no apreciaría nada extraño en estas unidades de medida.

El alumnado del nivel C debe evaluar si el modelo lineal es adecuado y si es un buen ajuste. El gráfico de residuos muestra que no hay patrones: hay cantidades parecidas de residuos negativos y positivos, y son tanto negativos como positivos a lo largo de los valores de x. Esto respalda la impresión inicial del diagrama de dispersión, que el modelo lineal es apropiado. Anteriormente se había indicado que el coeficiente de correlación era alto (0,8), lo que indica que la relación lineal es fuerte.

La pendiente de la recta de mínimos cuadrados (alrededor de 2,8) puede interpretarse como una estimación de la diferencia promedio en las estaturas de dos personas cuya longitud de los antebrazos difiere en 1 cm. Es decir, si la longitud del antebrazo de un miembro del coro es 1 cm mayor que la de otro, se espera que necesite una túnica 2,8 cm más larga. El punto de corte con el eje Y en 45,8 cm puede interpretarse teóricamente como la estatura esperada para una persona con un antebrazo de longitud 0 cm. Pero esto es claramente absurdo, por lo que, en este contexto, no tiene sentido interpretar la ordenada en el origen de la recta de mínimos cuadrados. No obstante, esta recta puede usarse razonablemente para predecir la estatura de una persona cuya longitud del antebrazo sea conocida, siempre que esa longitud esté en el rango de los datos empleados para construir la recta de regresión (de 39 a 50 cm, para estos datos). Por ejemplo, la estatura esperada para alguien con una longitud del antebrazo de 42 cm sería:

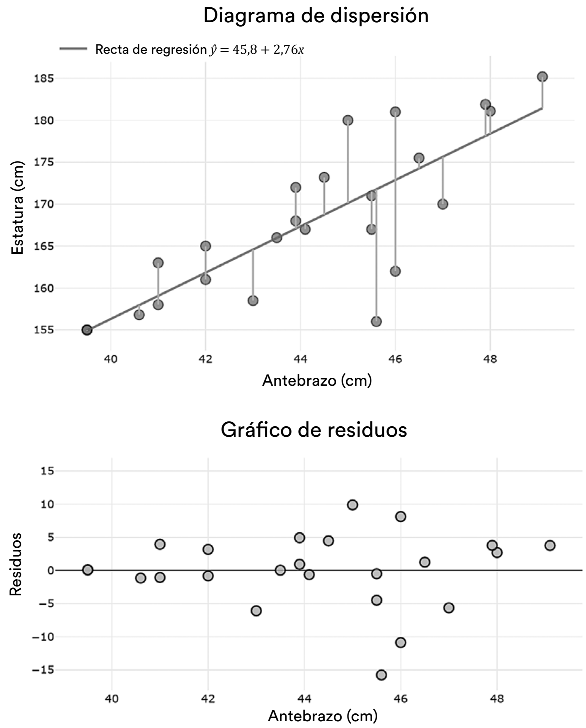


Figura 60. Recta de mínimos cuadrados y gráfico de residuos

$$\text{estatura esperada} = 45{,}8 + 2{,}76\,(42) = 161{,}7 \text{ cm}$$

Cuando se usa un modelo de ajuste para predecir un valor de y (la variable de respuesta) a partir de x (la variable explicativa), el intervalo de predicción asociado depende de la desviación típica de los residuos. Esto no debe confundirse con el intervalo de confianza para la respuesta media. La desviación típica de los residuos puede usarse para construir un intervalo de predicción para la estatura de todo el alumnado del centro con una longitud del antebrazo de 42 cm. Así, para un estudiante del centro escogido al azar con una longitud del antebrazo de 42 cm, se predice que tendrá una estatura en el rango de 150,1 cm a 173,3 cm. Este intervalo de predicción se calcula sumando y restando 2 · (desviación típica de los residuos) al valor de la estatura esperada. Al hacerlo así, se asume que la distribución de las estaturas del alumnado del centro con una longitud del antebrazo de 42 cm tiene forma de campana. Para estos datos, la desviación típica de los residuos es 5,8 (no se muestra aquí, pero forma parte de los resultados obtenidos por ordenador), así que 2 · (desviación típica de los residuos) = 2 · (5,8) = 11,6 cm. Sumar y restar esta cantidad a la estatura esperada de 161,7 cm proporciona un intervalo de predicción para la estatura de una persona con una longitud del antebrazo de 42 cm: de 150,1 cm a 173,3 cm.

Interpretar los resultados

El alumnado de la clase de Estadística concluye que la longitud del antebrazo y la estatura están relacionadas linealmente, y que la longitud del antebrazo puede usarse para predecir la estatura. Por ejemplo, para una longitud del antebrazo de 42 cm, se puede estar razonablemente seguro de que la estatura de cada uno de los estudiantes del centro estará entre 149,5 cm y 174,2 cm. Este intervalo es bastante amplio (un poco menos de 10 pulgadas),[26] lo que sugiere que, aunque el alumnado puede usar la ecuación de regresión para predecir la estatura, debe esperar una variación considerable y, por lo tanto, quizá necesite ajustar muchos dobladillos.

[26] N. de T.: 10 pulgadas son 25,4 cm, la amplitud del intervalo es de 24,7 cm.

Ejemplo 6: La siesta y los ataques al corazón. Inferir la asociación a partir de un estudio observacional

Formular preguntas de investigación estadística

El efecto de las siestas (hacer siesta o no) sobre las enfermedades cardiovasculares (ECV), como los ataques al corazón, es incierto. Algunos estudios indican que dormir la siesta tiene un efecto positivo; algunos, un efecto negativo; y otros, que no tiene ningún efecto. Sin embargo, en muchos de estos estudios no se tuvo en cuenta la frecuencia de las siestas. Una pregunta de investigación estadística que considerar es:

¿Existe alguna relación entre la frecuencia de las siestas y la ECV?

Recoger/considerar los datos

Los estudios observacionales suelen ser la única opción cuando es prácticamente imposible o no es ético asignar aleatoriamente a los sujetos a los tratamientos. Estas situaciones son habituales al estudiar las causas de las enfermedades. En el diseño de un estudio que responda la pregunta de investigación estadística mencionada, no es viable asignar aleatoriamente a los sujetos a grupos de tratamiento y pedirles que no duerman ninguna siesta a la semana, que duerman exactamente 1 o 2 siestas a la semana o que duerman más siestas a la semana, y confiar en que respeten el número de siestas semanales indicado durante un periodo de observación largo.

Como alternativa, un grupo de investigación de Suiza (Häusler et al., 2019) diseñó un estudio observacional que seguía a 3462 sujetos suizos sin una historia previa de ECV. Se solicitó a los sujetos que hicieran un autoinforme de la frecuencia de las siestas que habitualmente hacían en una semana. A partir de la información proporcionada, el grupo de investigación clasificó a los sujetos en cuatro grupos en función de la variable explicativa: ninguna siesta, 1 o 2 a la semana, de 3 a 5 a la semana, 6 o 7 a la semana. La variable de respuesta era si el sujeto tenía o no tenía («Sí» y «No», respectivamente) un episodio de ECV durante los siguientes 5 años. La tabla 22 muestra el resumen de los recuentos de los sujetos para las categorías de las variables explicativa y respuesta al término del periodo de 5 años.

Tabla 22. Frecuencia de las siestas y episodios de ECV

	Frecuencia de las siestas				
Episodio de ECV	**Ninguna siesta**	**1-2 a la semana**	**3-5 a la semana**	**6-7 a la semana**	**TOTAL**
Sí	93	12	22	28	155
No	1921	655	389	342	3307
TOTAL	2014	667	411	370	3462

Analizar los datos

Entender la naturaleza de los recuentos de la tabla 22 es esencial para analizar los datos. Cada observación se registra en una de las celdas interiores que muestran frecuencias conjuntas (donde se solapan las categorías de las dos variables). Además, cada observación es una de las 3462 observaciones totales, así como una de las frecuencias marginales por fila y una de las frecuencias marginales por columna. Por ejemplo, un sujeto clasificado como «Ninguna siesta» y «No» en «Episodio de ECV» forma parte de la frecuencia marginal de 1921, de los 3307 sujetos que no han tenido episodios de ECV y de los 2014 sujetos que no duermen la siesta. El análisis de la tabla puede conducir al alumnado a examinar si hay o no una diferencia notable en el porcentaje de sujetos que tuvieron un episodio de ECV para cada categoría de frecuencia de siestas. Por ejemplo, de los sujetos que no durmieron ninguna siesta a la semana, el porcentaje que tuvo un episodio de ECV fue 93/2014 = 4,6 %. De los sujetos que durmieron 1 o 2 siestas a la semana, el porcentaje que tuvo un episodio de ECV fue 12/667 = 1,8 %. La tabla 23 muestra todos estos porcentajes condicionados.

Tabla 23. Porcentajes condicionados de episodios de ECV para cada frecuencia de siestas

	Frecuencia de las siestas				
Episodio de ECV	**Ninguna siesta**	**1-2 a la semana**	**3-5 a la semana**	**6-7 a la semana**	**TOTAL**
Sí	4,6 %	1,8 %	5,4 %	7,6 %	4,6 %
No	94,4 %	98,2 %	94,7 %	92,4 %	94,4 %
TOTAL	100 %	100 %	100 %	100 %	100 %

Los distintos grupos de siestas presentan diferencias en los porcentajes de incidencia de un episodio de ECV. El grupo de 1 o 2 a la semana tuvo el menor porcentaje de incidencias: un 1,8 %, que es al menos 3 puntos porcentuales más bajo que el de los otros tres grupos. El grupo que no durmió ninguna siesta tuvo un porcentaje de incidencia del 4,6 %. El grupo de sujetos que durmió 6 o 7 siestas a la semana tuvo el porcentaje más alto de episodios de ECV: un 7,6 %. Estos resultados se muestran en el gráfico de barras porcentuales de la figura 61.

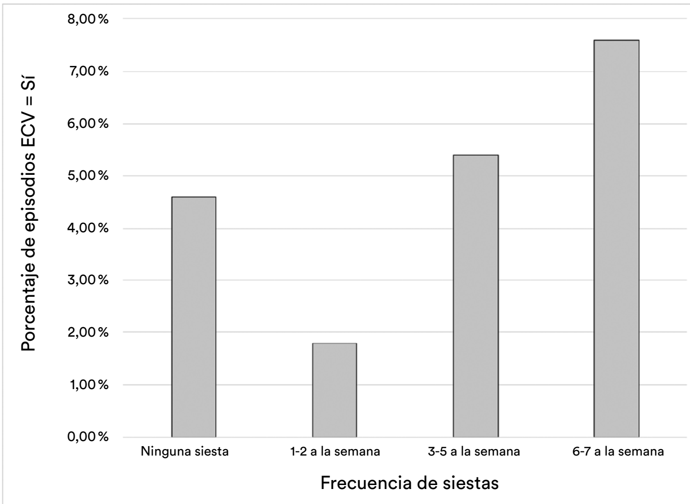


Figura 61. Porcentajes condicionados de episodios de ECV = Sí

Descriptivamente, parece que quienes no duermen ninguna siesta tienen un riesgo moderado de ECV a lo largo de los siguientes 5 años, quienes duermen 1 o 2 siestas semanales tienen el riesgo más bajo, y el riesgo se incrementa a medida que la persona duerme más siestas (estamos usando el porcentaje de tener una ECV condicionada por el número de siestas como medida del riesgo). Pero ¿podría deberse esta asociación meramente al azar? En otras palabras, ¿podría no haber ninguna relación entre el número de siestas y la aparición de una ECV?

Tabla 24. Valores esperados según los porcentajes condicionados

	Frecuencia de las siestas				
Episodio de ECV	**Ninguna siesta**	**1-2 a la semana**	**3-5 a la semana**	**6-7 a la semana**	**TOTAL**
Sí	90,17	29,86	18,40	16,57	155
No	1923,83	637,14	392,60	353,42	3307
TOTAL	2014	667	411	370	3462

Una estrategia es un test chi-cuadrado (χ^2), que compara el número de recuentos observados en cada celda con el número de recuentos esperados en cada celda si no hubiera ninguna asociación entre ambas variables (véase la tabla 24). Aunque la comprensión de los fundamentos técnicos de los test chi-cuadrado está más allá del nivel C, el alumnado puede usar herramientas tecnológicas e interpretar los resultados en un contexto. Los valores esperados pueden calcularse mediante los porcentajes marginales (ECV Sí = 155/3462 = 4,48 %; ECV No = 3307/3462 = 95,2 %), multiplicándolos por los totales de la columna correspondiente para obtener el siguiente resultado:

El estadístico chi-cuadrado es 20,3 con un p-valor de 0,0001 y 3 grados de libertad.

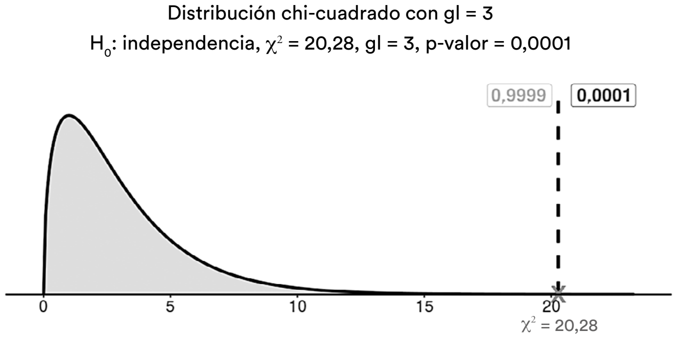


Figura 62. Resultados del test chi-cuadrado obtenidos mediante un *applet*

Interpretar los resultados

El test chi-cuadrado dio como resultado un p-valor de 0,0001 (véase la figura 62). Hay una probabilidad del 0,01 % de obtener un valor de 20,3, o mayor del estadístico del test chi-cuadrado si lo que se observa se debe únicamente al azar, bajo la hipótesis de que no hay ninguna relación. Por tanto, hay evidencia sólida para concluir que existe una relación entre el número de siestas realizadas y la aparición de una ECV. Nótese que la estrategia del chi-cuadrado no proporciona información acerca de la dirección de la asociación; no revela si un mayor número de siestas se asocia con un riesgo mayor o menor. De hecho, si se considera el gráfico de barras, la asociación podría ser no lineal. El riesgo es más alto para quienes no duermen ninguna siesta que para quienes duermen 1 o 2 a la semana, y se incrementa con el número de siestas. El riesgo para el grupo que no duerme ninguna siesta en comparación con el grupo que duerme 1 o 2 siestas a la semana podría medirse mediante el riesgo relativo (RR), calculado como RR = 4,6 % / 1,8 % = 2,55. Podemos interpretar este resultado como que el porcentaje de episodios de ECV es 2,55 veces mayor para quienes no duermen siestas que para quienes duermen 1 o 2 siestas a la semana.

¿Este estudio implica que todo el mundo debería empezar a dormir 1 o 2 siestas a la semana para evitar episodios de ECV y no superar este número de siestas? No necesariamente, ya que este estudio ha sido observacional y no puede hacerse una afirmación de causa-efecto. Podría haber variables de confusión,[27] más allá de la frecuencia de las siestas, que tengan un impacto en la variable respuesta de los episodios de ECV. Por ejemplo, si las personas mayores tienen más tendencia a dormir 6 o 7 siestas a la semana y también tienen más tendencia a tener una ECV, entonces podría ser importante considerar la variable edad. En este caso, habría una asociación entre la frecuencia de las siestas y la ECV incluso si las siestas no tuvieran ningún efecto en la ECV. De igual forma, si las personas con otros problemas de salud suelen dormir más siestas y tienen una mayor probabilidad de tener una ECV, entonces tener otros problemas de salud sería una variable de confusión en la relación con la frecuencia de las siestas.

Cuando no se pueden llevar a cabo experimentos y, en su lugar, se emplean estudios observacionales, es importante identificar posibles variables de confusión. En ese caso, si es posible, se deben controlar esas variables. Una forma de hacerlo es desagregar los datos para analizar las variables explicativa y de respuesta a través de las categorías de las variables de confusión. Cuando el alumnado progrese más allá del nivel C, aprenderá más sobre cómo controlar las posibles variables de confusión en el diseño del estudio. También aprenderá métodos estadísticos más formales para analizar los datos, como los intervalos de confianza para las diferencias, y la manera de evaluar la importancia práctica de la asociación mediante el análisis de los tamaños del efecto.

Ejemplo 7: La población en edad laboral. Trabajar con datos secundarios

Formular preguntas de investigación estadística

El alumnado y el público en general se encuentran habitualmente con gráficos estadísticos que rara vez se ven en las aulas de Pre-K–12. Una parte de la alfabetización estadística consiste en ser capaz de enfrentarse a los datos presentados en los medios a través de representaciones poco convencionales.

[27] N. de T.: En el original se habla de *confounding variables*. En la literatura en español se mencionan como *variables* o *factores de confusión* (sobre todo, en el ámbito de investigación sanitaria) y también como *variables extrañas* o *no controladas*.

El *New York Times* tiene una sección semanal, «What's Going On In This Graph?» (WGOITG; https://www.nytimes.com/column/whats-going-on-in-this-graph), que proporciona ejemplos excelentes de tales representaciones. Cada semana se presenta un gráfico estadístico que apareció en una noticia anterior, pero se elimina parte del contexto. Se invita al alumnado a publicar sus respuestas a las tres preguntas que se plantean: ¿qué observas?, ¿qué te preguntas? y ¿qué está pasando en este gráfico? Alternativamente, se puede pedir al alumnado que escriba un titular llamativo que plasme la idea principal del gráfico.

Por ejemplo, un WGOITG (The Learning Network, 2019) trataba el tema de los cambios de tamaño en la población en edad laboral a lo largo de un periodo de diez años. En un contexto como este, el objetivo principal es plantear una pregunta de investigación estadística que pueda responderse. Tras examinar este gráfico en concreto, el alumnado podría plantear la siguiente pregunta de investigación estadística:

¿Cómo ha cambiado la población en edad laboral en los Estados Unidos?

Recoger/considerar los datos

El gráfico proporcionado en el WGOITG (véase la figura 63) es un mapa de calor,[28] que representa un análisis estadístico de los datos. La leyenda facilitada nos ayuda a comprender los datos. El gráfico utiliza un código de colores para cada condado de los Estados Unidos de acuerdo con el cambio experimentado por la población en edad laboral. En este informe se presenta en una escala de grises. Por ejemplo, un condado está de color gris oscuro si su población en edad laboral (definida como de 25 a 54 años) ha disminuido más de un 10 % de 2007 a 2017. En el otro extremo de la gama de grises, el gris claro representa un condado cuya población en edad laboral se ha incrementado más de un 10 % de 2007 a 2017.

Cambio en la población en edad laboral (de 25 a 54 años) de 2007 a 2017

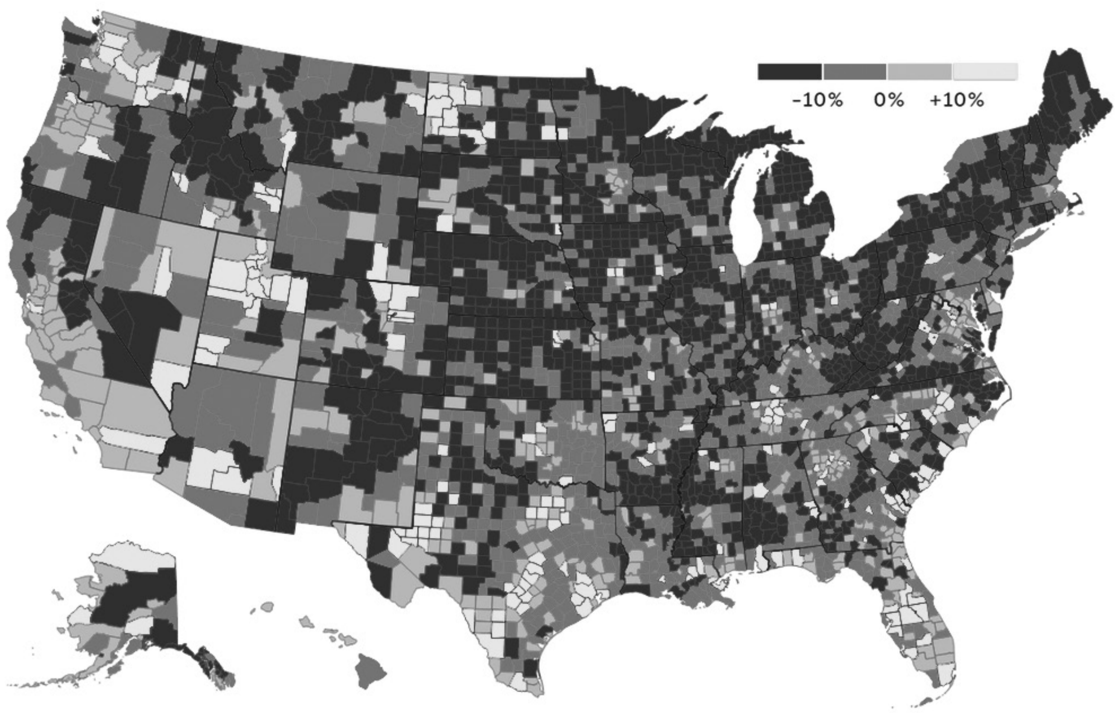


Figura 63. Mapa de calor de los Estados Unidos que muestra el cambio en la población (The Learning Network, 2019)

El alumnado del nivel C debe estudiar los datos secundarios resumidos que se proporcionan en el gráfico, para comprender las distintas variables que se muestran. Debe ser capaz de identificar las unidades observacionales (condados) y todas las variables representadas en este gráfico. Por ejemplo, una de esas variables es el

[28] N. de T.: En español también se usa a menudo la denominación en inglés, *heat map*. Este tipo de representación no debe confundirse con la coropleta: aunque ambos tipos de representación usan escalas de colores para representar la intensidad, muestran diferencias en relación con la manera en que se definen las áreas coloreadas.

cambio porcentual en el número de personas con edades comprendidas entre los 25 y los 54 años que residen en un condado. El conjunto de datos que produjo estos gráficos quizá tuviera columnas etiquetadas como «condado», «población_edad_laboral_2007», «población_edad_laboral_2017», «cambio_porcentual».

Analizar los datos

En este caso, ya se ha llevado a cabo el análisis. El propio gráfico es el análisis de los datos. Esto es habitual cuando se interpreta un trabajo ajeno tal y como aparece en medios de comunicación o en publicaciones científicas.

Interpretar los resultados

El alumnado del nivel C puede observar que en el centro y el nordeste del país predomina el color gris oscuro, lo que apunta a una disminución de la población en edad laboral en esas zonas, mientras que en el oeste y el sur hay una tendencia a aumentar la población en edad laboral. En general, los datos sugieren grandes cambios a lo largo del tiempo en el número de individuos en edad laboral. El alumnado también puede comentar la variabilidad dentro de cada estado. Algunos estados, como Nebraska, Maine o Vermont, están casi por completo de gris oscuro, lo que indica que todo el estado sufre un decrecimiento de la población en edad laboral. Otros estados, como Utah, están mayormente de gris claro, lo que representa grandes aumentos de la población en edad laboral. Otros, como Texas o California, tienen más variabilidad; algunas áreas del estado tienen aumentos y otras tienen disminuciones.

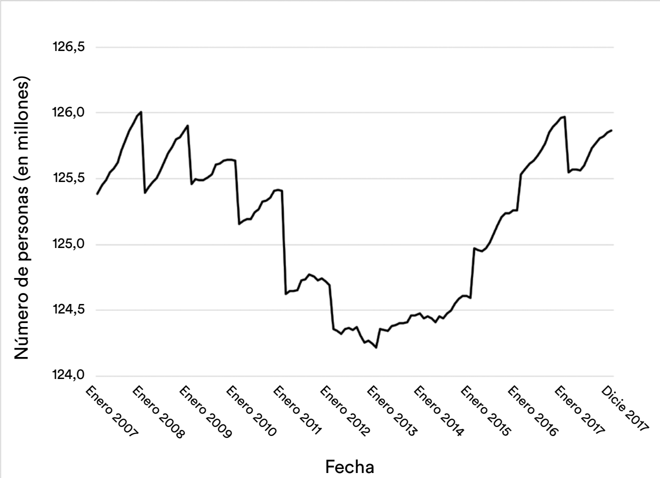


Figura 64. Gráfico temporal de los datos mensuales del mapa de calor

El alumnado del nivel C también puede consultar otras fuentes de información como ayuda para interpretar sus resultados. El mapa de calor del WGOITG mostrado anteriormente plasma la diferencia a lo largo de una década sin considerar los incrementos mensuales o anuales. Considerar los datos mensuales proporciona una oportunidad para examinar los datos más a fondo. El gráfico temporal (véaase la figura 64) muestra el total de individuos en edad laboral de los Estados Unidos correspondiente a los meses entre enero de 2007 y diciembre de 2017. La población en edad laboral disminuyó de manera general hasta 2014 y, a partir de 2015, parece que fue en aumento.

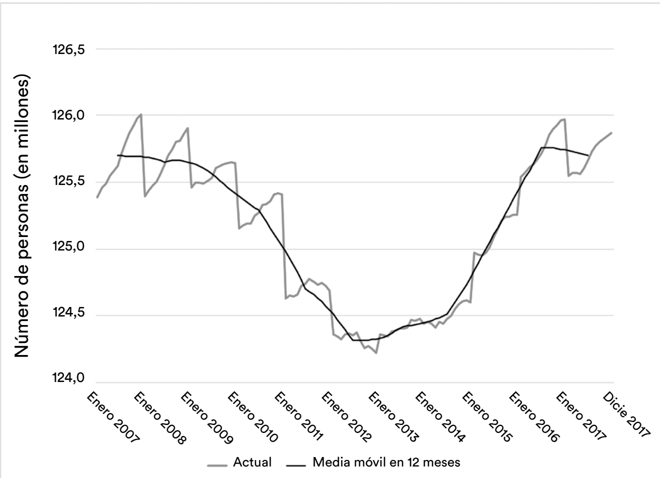


Figura 65. Gráfico temporal con la media móvil

El alumnado del nivel C puede observar el patrón del gráfico temporal, que presenta una caída desde el final de un año hasta el comienzo del siguiente (p. ej., el cómputo en diciembre de 2007 es de alrededor de 126 millones, pero cae en enero de 2008 a menos de 125,5 millones). Asimismo, podría suponer que la edad de cada persona se calculó en función del año de nacimiento en lugar del mes de nacimiento. Para tener esto en cuenta sin perder el nivel de detalle disponible en los datos mensuales, puede calcular una media móvil. Las medias móviles pueden variar en el periodo de tiempo empleado (p. ej., 3 meses, 6 meses, etc.); en este caso, tiene sentido usar una media móvil en 12 meses, lo que suavizará estos ajustes anuales. El gráfico de la figura 65 incluye los datos mensuales, así como la media móvil en 12 meses.

Si se dispone de los datos mensuales de 2007 a 2017, la primera media móvil en 12 meses que se puede calcular es la de junio de 2007 (engloba de enero de 2007 a diciembre de 2007), y el último valor que se puede obtener es el de junio de 2017 (comprende de enero de 2017 a diciembre de 2017). La tendencia de la media móvil es similar, pero más suave. Hay otras formas de suavizar los patrones de los gráficos, como los ajustes estacionales o el suavizado exponencial.

Ejemplo 8: La clasificación de lagartijas. Predecir una variable cualitativa

Formular preguntas de investigación estadística

El alumnado del nivel C va más allá que el del nivel B porque aborda preguntas más abiertas y considera conjuntos de datos que pueden requerir cierta preparación antes del análisis. Los conjuntos de datos apropiados para el nivel C pueden contener variables adicionales que no se usarán en el análisis, por lo que el alumnado deberá considerar diferentes aspectos de los datos y cómo se pueden seguir diferentes caminos de manera productiva.

Supongamos que, en una clase de Ciencias, se está explorando el impacto del desarrollo humano en la fauna salvaje. En un análisis anterior, el alumnado descubrió que las lagartijas de hábitats «alterados» (hábitats con un desarrollo humano considerable) tendían a tener mayor masa[29] que las de hábitats naturales. La clase planteó esta pregunta de investigación estadística:

¿Puede usarse la masa de una lagartija para predecir si proviene de un hábitat alterado o natural?

El alumnado puede afrontar esta pregunta de investigación estadística tratando de definir reglas sobre cómo clasificar una lagartija seleccionada aleatoriamente como proveniente de un hábitat alterado o natural.

Recoger/considerar los datos

Una bióloga capturó varios individuos de una especie de lagartija, *Anolis sagrei*, en dos tipos de hábitats diferentes en cada una de las cuatro islas de las Bahamas. Los datos fueron recogidos por Erin Marnocha durante su etapa como estudiante de posgrado en la Universidad de California, en Los Ángeles. Las lagartijas se seleccionaron mediante un proceso de muestreo aleatorio. Algunas vivían en hábitats desarrollados por humanos (alterados), mientras que otras vivían en hábitats naturales sin desarrollo humano (naturales).

Se capturó un total de 160 lagartijas, 81 de hábitats naturales y 79 de hábitats alterados. Una vez capturadas, se midieron varias de sus características físicas, incluyendo la masa, la longitud, la anchura, la anchura entre las patas, el ancho de la cabeza, etc. Las lagartijas capturadas en hábitats con desarrollo humano se clasificaron como «Alterado»; y las capturadas en hábitats sin desarrollo, como «Natural».

En la figura 66 se proporcionan las primeras filas de datos. Nótese que estos datos se muestran en formato expandido, de modo que cada fila representa una lagartija, y las columnas nos indican las características medidas.

Para considerar y comprender estos datos secundarios, el alumnado del nivel C puede examinar y cuestionar el conjunto de datos mediante herramientas tecnológicas. Debe darse cuenta de que, aunque el conjunto de datos incluye 160 lagartijas, unas cuantas tienen valores ausentes para una u otra variable. A la variable «Longitud de la cola» le falta un número considerable de datos (p. ej., índice 3 de la figura 66). Quienes hayan cazado lagartijas sabrán por qué: las lagartijas pueden perder la cola para evitar que las atrapen.

[29] N. de T.: En el original se utiliza el término *mass* para referirse a la cantidad de materia que contiene un cuerpo; por este motivo, se ha optado por mantener este término en la traducción al español, en vez de usar el término *peso*, que se refiere a la fuerza de gravitación universal que ejerce un cuerpo celeste sobre una masa.

Lagartijas de Marnocha												
Casos (160 casos)												
Índice	Isla	Hábitat	Masa (g)	LHC (mm)	Longitud de la extremidad posterior (mm)	Anchura entre extremidades posteriores (mm)	Longitud de la extremidad anterior (mm)	Anchura entre extremidades anteriores (mm)	Anchura de la apertura bucal (mm)	Profundidad de la cabeza (mm)	Anchura de la almohadilla de los dedos (mm)	Longitud de la cola (mm)
1	New Pr...	Natural	2	49	11,2	23,3	7,1	20,5	7	4	0,9	
2	New Pr...	Natural	2,2	45	10,4	25,4	7	20,8	7,6	5,1	1,2	78
3	New Pr...	Natural	2,4	50	11,6	27,1	7,1	22,5	7,2	4,6	1	
4	New Pr...	Natural	2,6	48	11,2	25,8	7,2	21,4	7,8	4,5	1,1	83
5	Exuma	Natural	2,7	53	11,4	27	7,3	22,9	7	5,1	0,8	
6	New Pr...	Natural	2,7	48	11,3	25,1	8,1	21,9	7,2	4,5	1,2	
7	Harbor I...	Natural	2,7	47	10,8	24,2	7	19,9	8,5	4,7	0,9	83
8	Eleuthe...	Natural	2,7	46	10,9	25	7,3	21	7,5	4,6	1,1	
9	Exuma	Natural	2,8	49	12,2	25,8	7,2	22,6	7,4	4,4	1,1	94
10	New Pr...	Natural	2,8	49	11,9	25,7	8,7	23,2	8	4,5	0,9	89
11	New Pr...	Natural	2,8	51	11,4	25,3	7,6	22,7	8,1	4,8	1,1	
12	New Pr...	Natural	2,8	50	12,1	26,4	7,6	22,5	7,4	4,4	1,2	
13	New Pr...	Natural	2,9	51	11,8	27	8,8	23,3	8,7	5,2	1,1	90
14	Eleuthe...	Natural	2,9	48	11,5	25,8	7	21,6	8,1	4,6	1	
15	New Pr...	Natural	3	51	10,4	26,2	8,1	22,5	7,3	4,3	1,1	86
16	New Pr...	Alterado	3	52	11,6	25	7,5	21	7,7	4,1	0,9	

Figura 66. Primeras filas de datos de las lagartijas en formato extendido

Esta pregunta de investigación estadística depende solo de dos variables: «Hábitat» y «Masa» (medida en gramos). Para facilitar el análisis, el alumnado puede crear un nuevo conjunto de datos que contenga solo esas dos variables.

Analizar los datos

El gráfico de puntos comparativo de la masa (véase la figura 67) muestra un solapamiento considerable entre los dos tipos de hábitats. La pregunta de investigación estadística se centra en predecir si una lagartija seleccionada al azar proviene de un hábitat alterado o natural (variable cualitativa de respuesta), utilizando, para ello, la masa de la lagartija como la variable explicativa. En el nivel C, el alumnado puede abordar el análisis mediante la clasificación (más allá del nivel C, hay otros métodos de análisis que pueden ayudar a responder este tipo de preguntas de investigación estadística).

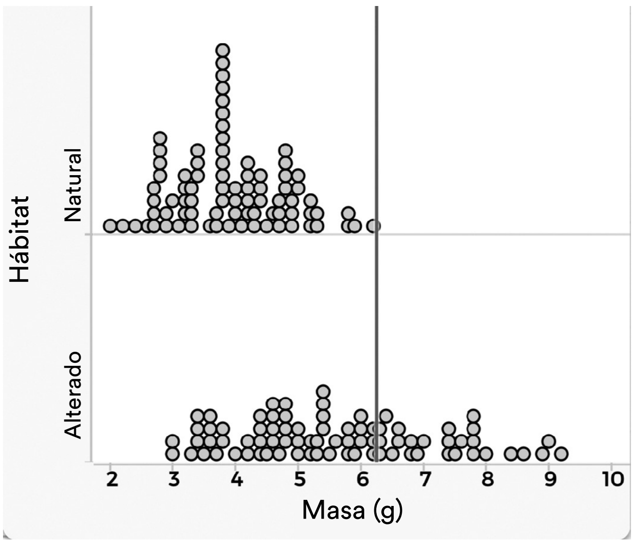


Figura 67. Gráficos de puntos apilados con una masa de 6,25 g como punto de corte

Esencialmente, un método de clasificación requiere que se proponga, a partir de la masa de la lagartija, una regla que prediga el tipo de ambiente del que proviene. Por ejemplo, consideremos una posible regla que clasifique las lagartijas de menos de 6,25 g como provenientes de hábitats naturales («Natural»). El alumnado del nivel C puede analizar si esta posible regla es o no es una buena regla de predicción.

El gráfico de puntos revela que, si se clasifican las lagartijas que pesan menos de 6,25 g como provenientes de hábitats naturales, entonces el 100 % de estas lagartijas están correctamente clasificadas (porque todas estas lagartijas en nuestra muestra pesan menos de 6,25 g). Sin embargo, esta regla de clasificación no es perfecta, porque un alto porcentaje (53/79 = 0,671 = 67,1 %) de las lagartijas provenientes de hábitats alterados serían incorrectamente clasificadas, ya que pesan menos de 6,25 g. Dado que el gráfico de puntos muestra mucho solapamiento entre las masas de ambos grupos, la clasificación de una lagartija como proveniente de un hábitat alterado o natural no será inmediata.

El alumnado del nivel C debe saber que un menor solapamiento entre las distribuciones da lugar a una mayor tasa de éxito en la clasificación. Además, cambiar el punto de corte puede hacer que la tasa de clasificaciones erróneas varíe. Por ejemplo, si las lagartijas con masas inferiores a 5,0 g se clasificaran como «Natural», se cometerían menos errores al etiquetar a lagartijas que provengan de hábitats alterados, pero se cometerían más errores con lagartijas que realmente provengan de hábitats naturales.

El análisis de los datos mediante el método de la clasificación está al alcance del alumnado del nivel C. Con la tecnología apropiada o mediante el recuento riguroso de puntos en el gráfico, puede calcular e interpretar la matriz de confusión para cualquier regla de punto de corte que elija. Una matriz de confusión presenta los valores reales en las columnas y las categorías de clasificación en las filas (o viceversa). Por ejemplo, en la celda superior izquierda de la matriz de la tabla 25 se indica que hubo 26 lagartijas correctamente clasificadas como provenientes de un hábitat alterado. Un clasificador perfecto tendría ceros en todas las celdas fuera de la diagonal.

La primera matriz de confusión de la tabla 25 muestra los resultados de la clasificación con la regla según la cual las lagartijas con masas inferiores a 6,25 g se clasifican como «Natural». Las 81 lagartijas provenientes de hábitats naturales fueron clasificadas como «Natural», pero 53/79 lagartijas de hábitats alterados fueron clasificadas incorrectamente. En conjunto, la tasa de clasificación errónea es (número de lagartijas mal clasificadas) / (número total de lagartijas) = (53 + 0) / (26 + 53 + 0 + 81) = 0,331, o aproximadamente 33 %.

Tabla 25. Matriz de confusión para masa < 6,25 g

	Realmente provenientes de un hábitat alterado	Realmente provenientes de un hábitat natural
Clasificadas como «Alterado»	26	0
Clasificadas como «Natural»	53	81

Si se modifica la regla para clasificar una lagartija como «Natural» cuando su masa es inferior a 5 g, se obtiene una nueva matriz de confusión:

Tabla 26. Matriz de confusión para masa < 5 g

	Realmente provenientes de un hábitat alterado	Realmente provenientes de un hábitat natural
Clasificadas como «Alterado»	49	11
Clasificadas como «Natural»	30	70

El alumnado puede seguir analizando reglas de clasificación para determinar cuál produce la tasa de clasificación errónea más baja.

Determinar un procedimiento óptimo para la clasificación queda fuera del alcance del nivel C. En este nivel, el alumnado debe comprender, mediante el cálculo de probabilidades, cómo afectan los cambios de una regla a las tasas de clasificación errónea. Asimismo, debe entender que se puede utilizar un método de clasificación para responder preguntas de investigación estadística centradas en predecir una variable de respuesta cualitativa a partir de otra variable. Superado el nivel C, hay otros métodos para responder esas preguntas (p. ej., la regresión logística, los bosques aleatorios,[30] el aprendizaje automático[31] y los algoritmos de aprendizaje profundo).[32] En general, la clasificación forma parte de la predicción; aquí, por ejemplo, se intenta predecir la categoría de una lagartija escogida al azar a partir de una o más variables. Estos métodos

[30] N. de T.: Con frecuencia referidos en español por el término en inglés, *random forests*.
[31] N. de T.: También se utiliza el término inglés *machine learning*.
[32] N. de T.: A menudo denominado con el término inglés *deep learning*.

aparecen con frecuencia en las noticias, especialmente en relación con los avances en aprendizaje automático o profundo. El reconocimiento facial o de imágenes es un ejemplo avanzado de un problema de clasificación abordado con frecuencia por los medios actualmente (p. ej., Singer y Metz, 2019). Para fomentar la alfabetización estadística, en el nivel C, el alumnado debe aprender las ideas básicas de los métodos de clasificación.

En el nivel C, a partir de ciertos datos sobre una nueva lagartija de la población, el alumnado también debe ser capaz de clasificarla y de proporcionar la probabilidad de clasificación errónea para esa nueva observación.

Interpretar los resultados

Una tasa de clasificación errónea del 26 % podría parecerle buena a parte de la clase y no tan buena a la otra parte. El alumnado podría preguntarse cómo evaluar estas tasas. Para dar con una tasa «de referencia» para hacer comparaciones, un enfoque es ignorar la masa de las lagartijas y clasificarlas todas en el grupo de mayor tamaño. Dado que hay 79 lagartijas de hábitats alterados y 81 de hábitats naturales, se pueden clasificar todas como «Natural», sin tener en cuenta su masa.

Esta regla clasificará incorrectamente todas las lagartijas de hábitats alterados, por lo que la tasa de clasificación errónea será 79 / (81 + 79) = 0,49 o 49 %. En comparación, la regla de clasificación anterior con una tasa de error del 26 % (las lagartijas con una masa inferior a 5,0 g son de hábitats naturales) es bastante mejor que esta tasa de referencia.

Dado que las lagartijas se muestrearon aleatoriamente, se pueden interpretar estas tasas de clasificación errónea como estimaciones de la proporción de clasificaciones incorrectas para futuras lagartijas capturadas, siempre que se hayan encontrado en la misma población que las lagartijas de este conjunto de datos. El alumnado del nivel C puede concluir que, de hecho, las lagartijas pueden clasificarse según su masa de manera que la probabilidad de clasificación errónea sea menor que si, simplemente, se clasificaran todas como provenientes del mismo hábitat.

Revisión del análisis

El enfoque anterior depende de una única variable explicativa: la masa de la lagartija. ¿Cómo podrían usarse múltiples variables para clasificar las lagartijas? Más específicamente, la pregunta de investigación estadística podría reformularse de la siguiente forma:

¿Es posible usar la masa de una lagartija, la profundidad de su cabeza y la envergadura de sus extremidades posteriores para predecir si proviene de un hábitat alterado o natural?

Un método que va más allá de la simple clasificación y que resulta accesible para el alumnado del nivel C si dispone de la tecnología adecuada es la aplicación de un árbol de clasificación y regresión (CART, por su sigla en inglés).[33] El CART es un ejemplo de método estadístico actual que se apoya en un algoritmo más que en un modelo matemático. En este caso, se entiende por «algoritmo» una serie de reglas.

Aunque la aplicación e interpretación del algoritmo CART es accesible para el alumnado del nivel C, no es necesario que conozca los detalles para que empiece a entenderlo. CART comienza separando los datos de acuerdo con una regla. Supongamos que esa regla es la indicada anteriormente: se clasifica una lagartija como «Natural» si su masa es inferior a 5 g. Esta regla divide los datos en dos grupos.

Un grupo consiste en lagartijas de menos de 5 g que son clasificadas como «Natural» (incluso a pesar de que este grupo pueda contener algunas que no provengan de un hábitat natural), y el otro grupo lo conforman lagartijas que pesan más de 5 g y que son clasificadas como «Alterado» (aunque algunas puedan no proceder de un hábitat alterado).

El alumnado puede considerar entonces el grupo de lagartijas clasificadas como «Natural» y determinar si, basándose en una de las variables del conjunto de datos, puede dividir una vez más este grupo en otros dos grupos más pequeños.

[33] N. de T.: *Classification and regression tree.*

Por ejemplo, la división inicial clasificó 100 lagartijas como «Natural» porque su masa era inferior a 5 g (véase la matriz de confusión de la tabla 26, donde 30 + 70 = 100). Ahora, al considerar solo estas 100 lagartijas, el alumnado puede repetir el proceso con una nueva variable; por ejemplo, mediante la representación de gráficos de puntos apilados para las medidas de la profundidad de la cabeza de las 100 lagartijas. Los resultados se presentan en la figura 68: en la parte superior se muestran las lagartijas verdaderamente provenientes de un hábitat natural; y en la inferior, las provenientes de un hábitat alterado.

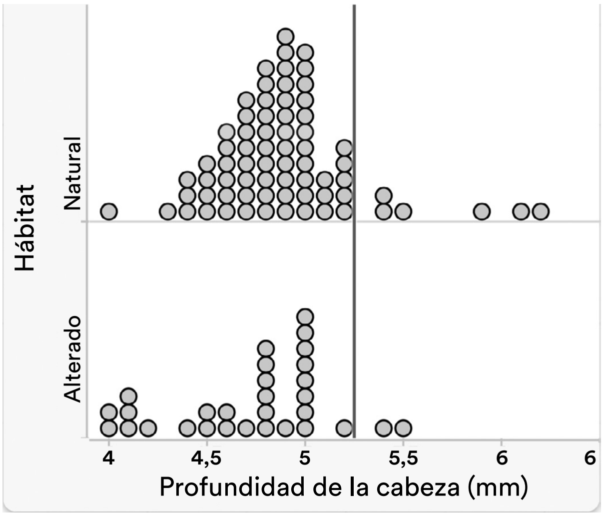


Figura 68. Gráficos de puntos apilados para 100 lagartijas con masas inferiores a 5 g; se establece el corte en una profundidad de la cabeza de 5,25 mm.

De nuevo, el alumnado puede considerar distintos valores «de corte». Un primer intento podría ser dibujar el límite en una profundidad de la cabeza de 5,25 mm, ya que así, visualmente, la mayor parte de los datos queda a la izquierda, separada del resto, a la derecha. Dado que las lagartijas de hábitats naturales tienen una media de profundidad de la cabeza ligeramente mayor, el alumnado puede clasificar las que están a la derecha de la línea de corte de la figura 68 como «Natural», y las que están a la izquierda como «Alterado». Esto conduce a un nuevo conjunto de reglas de clasificación:

1) *Si la masa es inferior a 5 g, se considera la profundidad de la cabeza. Si la profundidad de la cabeza es inferior a 5,25 mm, se clasifica como «Alterado»; y si es mayor o igual que 5,25 mm, se clasifica como «Natural».*
2) *Si la masa es igual o superior a 5 g, se clasifica como «Alterado».*

El siguiente paso sería repetir este procedimiento en las 60 lagartijas cuyas masas eran iguales o superiores a 5 g y que, por tanto, habían sido clasificadas como «Alterado».

El proceso puede continuar durante un buen rato. En cada paso se puede considerar cualquier variable para hacer la separación de los datos en dos grupos de clasificación. El algoritmo CART hace lo que ningún ser humano tendría la paciencia suficiente para hacer. En cada paso del proceso considera todas las variables disponibles y todos los posibles valores de corte. El CART determina cuál proporciona la tasa de clasificación errónea más baja y, después, separa los datos en función de esa regla. Así se crean dos nuevos grupos con los que se repite el proceso. Esto, claro, produce cuatro grupos, y el proceso vuelve a repetirse con cada uno de ellos. Se termina cuando no hay ninguna nueva división que mejore la tasa de clasificación errónea o cuando un grupo es demasiado pequeño para dividirlo.

El resultado puede escribirse como una serie de reglas, pero se aprecia mejor en forma de árbol. La figura 69 muestra el CART para estos datos. Cada división contiene una regla enunciada como una condición. Si la condición es cierta, nos movemos hacia la rama inferior izquierda, y así sucesivamente hasta alcanzar un nodo terminal. El nodo indica cómo clasificar la observación.

Por ejemplo, consideremos una lagartija con una masa de 4,5 g, una profundidad de la cabeza de 5 mm y una longitud de las extremidades posteriores de 11 mm. La primera regla (masa ≥ 5,35 g) no se verifica en esta lagartija, lo que indica que tenemos que movernos a la rama derecha. La siguiente regla considera la profundidad de la cabeza; dado que la profundidad de la cabeza de la lagartija no es inferior a 4,25 mm, nos movemos de nuevo hacia la derecha. A continuación, volvemos a considerar la masa, y puesto que es mayor que 4,35 g, esta vez el movimiento es hacia la izquierda. Finalmente, dado que la medida de las extremidades posteriores es inferior a 12,65 mm, el movimiento es hacia la izquierda, por lo que la lagartija se clasifica como «Alterado». Si la longitud de las extremidades posteriores hubiera sido superior a 12,65 mm, la lagartija se habría clasificado como «Natural».

Los números que figuran bajo las etiquetas de la clasificación indican el número de lagartijas de ambos tipos que se clasificaron con esa etiqueta. Por ejemplo, respecto al nodo terminal compartido por nuestra lagartija hipotética («Alterado 9 1»), 9 de las lagartijas de los datos muestrales que fueron enviadas a ese nodo procedían realmente de un hábitat alterado y 1 provenía de un hábitat natural. Por tanto, para lagartijas con características similares a estas, la tasa de clasificación errónea es de 1/10, o de un 10 %.

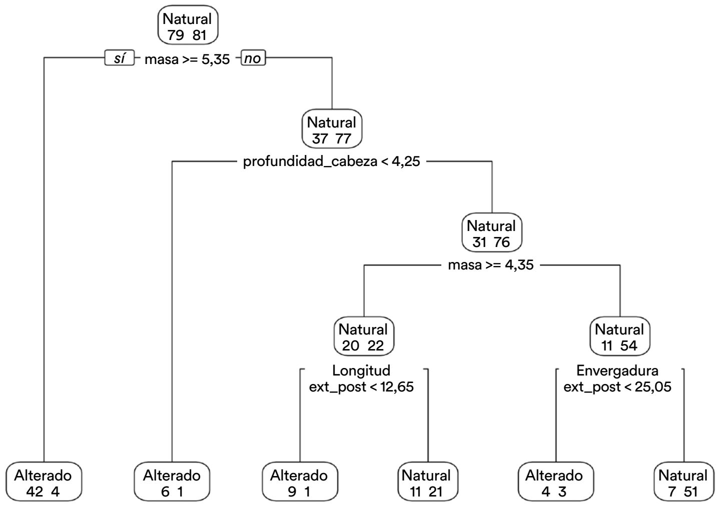


Figura 69. CART para los datos de las lagartijas en el que se toma la rama izquierda cuando se verifica la regla.

Interpretar los resultados

Cada división del árbol es una decisión, y se clasificará cada lagartija de acuerdo con los nodos terminales del árbol. De los dos números de cada nodo, el de la izquierda representa el número de lagartijas cuyo verdadero hábitat de procedencia era alterado; y el de la derecha, el número de lagartijas de hábitats realmente naturales. Por ejemplo, el nodo del extremo izquierdo clasifica las lagartijas como «Alterado». Esta clasificación es correcta para 42 lagartijas e incorrecta para 4.

La tasa de clasificación errónea general es 18 % (27 de 160 lagartijas), una cifra calculada por el *software* que elaboró el árbol. Podemos confirmarlo contando el número de clasificaciones erróneas y dividiéndolo entre 160, el número total de lagartijas. Por ejemplo, el nodo que está a la izquierda del todo, etiquetado como «Alterado», contiene 42 lagartijas de hábitats alterados y 4 de hábitats naturales, por lo que hay 4 clasificaciones erróneas. Si continuamos de izquierda a derecha, encontraremos 1, 1, 11, 3 y 7 clasificaciones erróneas más, lo que hace un total de 27 clasificaciones erróneas en 160 lagartijas. Esto proporciona una tasa de clasificación errónea de 27/160 o alrededor del 18 %. Este árbol mejora la regla de clasificación anterior, que se basaba únicamente en la masa.

Si no se dispone de herramientas tecnológicas para crear árboles, también es beneficioso para el alumnado crear árboles a partir de conjuntos de reglas de su propio diseño.

Resumen del nivel C

El alumnado del nivel C debe tener soltura en el uso de preguntas a lo largo del proceso de resolución de problemas estadísticos. En este nivel, las preguntas de investigación estadística van más allá de las situaciones aditivas y comparativas para incluir preguntas acerca de asociaciones y relaciones entre múltiples variables, incluyendo predicciones.

Una vez implementado un plan adecuado para recoger datos con los que responder una pregunta de investigación estadística, y disponiendo de los datos resultantes, el paso siguiente suele ser resumirlos mediante herramientas como representaciones gráficas y resúmenes numéricos.

En el nivel C, el alumnado debe ser capaz de seleccionar técnicas de análisis de datos apropiadas para el tipo de datos de los que dispone, producir análisis estadísticos descriptivos y reconocer en el contexto las características importantes de esos datos. También debe ser capaz de proporcionar interpretaciones más sofisticadas y profundas en comparación con los niveles A y B. Estas interpretaciones tienen que integrar el contexto y los objetivos de la investigación para extraer conclusiones a partir de los datos y para respaldar dichas conclusiones a través de la evidencia estadística (tanto descriptiva como inferencial).

El alumnado del nivel C debe considerar la estadística como un conjunto de herramientas poderosas que le permite responder preguntas de investigación estadística y tomar decisiones informadas y fiables. También deben comprender las limitaciones de las conclusiones basadas en los datos de encuestas muestrales y de experimentos, y ser capaces de cuantificar la incertidumbre asociada a estas conclusiones utilizando el margen de error y propiedades relacionadas con las distribuciones muestrales. Asimismo, el alumnado estadísticamente competente mantiene un «escepticismo sano» en relación con la información estadística y tiene la capacidad de plantear preguntas para evaluar la validez de los hallazgos estadísticos. Al mismo tiempo, es consciente del poder del análisis estadístico para encontrar información y significado en los datos. Por otra parte, debe entender que, dado que otras personas leerán sus trabajos, tiene que documentar cuidadosamente tanto sus datos como sus análisis para que se puedan reproducir. Igualmente, debe comprender las consecuencias éticas de sus experimentos y análisis.

Si bien el nivel C completa la alfabetización estadística en Pre-K–12, el alumnado debe saber que existen preguntas de investigación estadística, métodos de recogida de datos, diseños de estudios y métodos de análisis más complejos. Hay una gran cantidad de conocimientos estadísticos que aprender después del nivel C. El contenido de los niveles A, B y C proporciona las bases de la alfabetización estadística en este momento; sin embargo, continúa evolucionando. No solo cambian los datos, sino también los procesos, la tecnología, las filosofías, las estrategias y muchos otros aspectos que impulsan la evolución de los métodos estadísticos. La estadística no es una disciplina estancada; por el contrario, es un campo en evolución arraigado en el proceso de resolución de problemas estadísticos que se ha analizado en este informe.

Evaluación

Las evaluaciones proporcionan retroalimentación al alumnado, al profesorado y a las familias, ya que permiten conocer qué sabe y qué puede hacer cada estudiante en un momento determinado. Es fundamental que la enseñanza y la evaluación estén alineadas, en consonancia con los principios establecidos en los distintos niveles del GAISE II. La evaluación del pensamiento estadístico del alumnado debería valorar la comprensión conceptual, estar situada en un contexto y requerir interpretación (Peck et al., 2013). Estos criterios se aplican a la evaluación formativa y a la sumativa, con independencia de su formato (p. ej., debates, actividades, proyectos, ensayos reflexivos, tareas para casa o pruebas). Aunque puede que no sea factible evaluar el razonamiento estadístico en todas las etapas del proceso de resolución de problemas estadísticos en un único ítem, es importante construir ítems que evalúen las diferentes partes del proceso a lo largo del tiempo.

Muchos ítems estadísticos que se encuentran habitualmente en evaluaciones estandarizadas o realizadas localmente requieren habilidades cognitivas de nivel inferior (como la memorización) y se centran en procedimientos y definiciones. Además, muchos de estos ítems no están centrados en el razonamiento estadístico, sino en los cálculos matemáticos. Por ejemplo, una tarea que consista en encontrar la media de un conjunto de números es una tarea de bajo nivel que implica cálculos matemáticos y no razonamiento estadístico. Incluso una tarea en la que el alumnado, disponiendo de todos los datos excepto uno y la media, deba encontrar el dato faltante es una tarea matemática de nivel bajo. Afortunadamente, existen varios ejemplos de ítems de evaluación robustos en evaluaciones estandarizadas, así como recursos para que el profesorado encuentre este tipo de ítems para usarlos en el aula.

Evaluaciones estandarizadas nacionales[34] e internacionales

Muchas evaluaciones estandarizadas nacionales e internaciones abordan el análisis de datos y la estadística. Por ejemplo, el Programa para la Evaluación Internacional de los Estudiantes (PISA, por su sigla en inglés),[35] que evalúa las habilidades de aplicación de conocimiento a situaciones del mundo real por parte de estudiantes de 15 años de distintos países del mundo, se enfoca en las matemáticas, con las simulaciones y la toma de decisiones como temas en los que se hace énfasis. De acuerdo con el College Board, el 29 % de los ítems de matemáticas del SAT[36] abordan la resolución de problemas y el análisis de datos (lo que incluye el razonamiento proporcional y la probabilidad). El análisis de datos, la estadística y la probabilidad conforman uno de los cinco ejes de contenido evaluados en el National Assessment of Educational Progress (NAEP). Y, por supuesto, el examen Advanced Placement[37] (AP) de estadística solo contiene preguntas sobre estadística. El formato estándar del examen incluye cuarenta preguntas de opción múltiple, cinco problemas breves de respuesta libre y una tarea de investigación. A lo largo de los años, el enfoque del examen ha evolucionado para centrarse en la comprensión conceptual y en la interpretación, en lugar de limitarse al cálculo aritmético. Los ítems seleccionados se publican tras cada convocatoria y están disponibles para el profesorado.

Fuentes de ítems de calidad para docentes

Crear ítems de evaluación de alta calidad es una tarea que requiere mucho tiempo y que puede suponer un reto. Por este motivo, describimos aquí dos fuentes que han sido validadas y que pueden servir como

[34] N. de T.: Aquí «nacionales» hace referencia al contexto de los Estados Unidos.
[35] N. de T.: Programme for International Student Assessment.
[36] N. de T.: El SAT (Scholastic Assessment Test) es el examen estandarizado de admisión universitaria utilizado en los Estados Unidos y organizado por el College Board.
[37] N. de T.: El Advanced Placement es una evaluación de nivel universitario dirigida a estudiantes de secundaria en el sistema educativo estadounidense, realizada por el College Board.

recursos para el profesorado: LOCUS (Levels of Conceptual Understanding in Statistics; https://locus.statisticseducation.org/) y STEW (Statistics Education Web; ASA, s.f.-c).

LOCUS

Las evaluaciones de LOCUS se basan en el marco de trabajo original del GAISE I, están en la línea de los estándares básicos comunes estadounidenses (Common Core State Standards) y se han diseñado como medidas fiables de la comprensión a lo largo de los niveles de desarrollo y a través del proceso de resolución de problemas estadísticos. Incluyen preguntas tanto de opción múltiple como de respuesta abierta, y hay diferentes versiones de las evaluaciones en función de los distintos niveles GAISE. Cada versión tiene formularios equiparados para fines de pre/post-test, por lo que se pueden utilizar tanto para evaluaciones formativas como sumativas.

Ambos tipos de preguntas, de opción múltiple y de respuesta abierta, están organizados por curso y por componente del proceso de resolución de problemas estadísticos. Las preguntas de opción múltiple de ejemplo incluyen los estándares asociados, el desempeño del alumnado, las respuestas correctas y comentarios explicativos. Las preguntas de respuesta abierta incluyen información adicional sobre las rúbricas de calificación, los errores comunes, ejemplos de respuestas y recursos complementarios.

STEW

El recurso en línea STEW incluye unidades didácticas[38] revisadas por pares. Las unidades didácticas de STEW siguen un formato estándar e incluyen los objetivos, los estándares de Common Core aplicables y las instrucciones para desarrollar la unidad en el aula. Aunque están organizadas por curso, muchas unidades didácticas pueden modificarse para abordar contenidos similares en cursos diferentes. El material puede incluir también problemas con ejemplos de resoluciones, diapositivas que se pueden utilizar para guiar los debates en el aula, e instrucciones para usar herramientas tecnológicas. Animamos al profesorado a que utilice las unidades didácticas disponibles en esta página y a que contribuya con nuevas unidades que hayan resultado efectivas en su aula. Estas unidades se pueden adaptar a evaluaciones.

Algunos de los ejemplos de ítems de evaluación que se ofrecen a continuación provienen de los recursos mencionados (p. ej., LOCUS y STEW) y se presentan por nivel. Estos ejemplos no pretenden ser exhaustivos, sino una muestra de cómo se puede evaluar el pensamiento estadístico.

Ejemplos de evaluación del nivel A

Ejemplo 1: Medidas de tendencia central

El alumnado de la profesora Kieffer lleva un registro del número de personas que hicieron uso del comedor escolar durante el mes de febrero. En la tabla 27 se muestra lo que encontraron:

Tabla 27. Número de estudiantes que usan el comedor escolar

L	M	X	J	V
18	21	19	20	22
25	17	19	18	19
19	20	21	21	23
21	20	23	19	23

Si la directora te preguntara cuántas personas de tu clase, aproximadamente, comen en el comedor escolar cada día, ¿qué le dirías? Explica cómo has obtenido tu respuesta.

[38] N. de T. Se ha optado por traducir el término *lesson plan* como «unidad didáctica», puesto que es más común en el contexto español que la traducción literal «plan de lección».

Ejemplo 2: Variabilidad, muestreo e inferencias

Quiero saber cuál es la comida favorita de toda la población mundial. ¿Obtendría una respuesta razonable si usara tu clase como muestra? ¿Por qué sí o por qué no?

Ejemplo 3: Cómo modelar los datos de diferentes maneras para ver de qué modo varían las conclusiones en función de la representación de los datos

(de la unidad didáctica «Candy Judging» de la web STEW)

La clase de Jeremy clasificó las chocolatinas de la más preferida (1) a la menos preferida (4). Estos datos se muestran en la tabla 28.

Tabla 28. Clasificación de la más preferida (1) a la menos preferida (4)

	Special Dark®	krackel®	mr. Goodbar®	Hershey® Milk Chocolate
Jeremy	1	2	3	4
Kayla	4	2	3	1
Quentin	1	2	3	4
Ken	4	3	1	2
Jake	1	3	4	2
Polly Ann	1	2	3	4
Rocco	1	3	2	4
Drake	4	2	1	3
Corrine	1	2	3	4
Kris	4	2	1	3
Mary	4	3	2	1
Casey	1	3	4	2
Mel	4	3	2	1
Lisa	1	3	2	4
Cindy	4	2	1	3

Utiliza los siguientes métodos para determinar la chocolatina favorita de la clase:

(a) Encuentra la chocolatina que se escogió más veces como la favorita (clasificada con un 1).

Representa este dato en un pictograma (o un gráfico de puntos). ¿Qué chocolatina representaría la favorita de la clase si se usara este método para determinarla?

(b) Localiza la chocolatina que se escogió más veces como la menos favorita (clasificada con un 4).

Representa este dato en un pictograma (o un gráfico de puntos). ¿Qué chocolatina representaría la menos favorita de la clase si se usara este método para determinarla?

(c) Calcula la suma de las puntuaciones de las clasificaciones. ¿Qué chocolatina representaría la favorita si se determinara con la suma de las puntuaciones de toda la clase? ¿Qué chocolatina sería la menos favorita con este método?

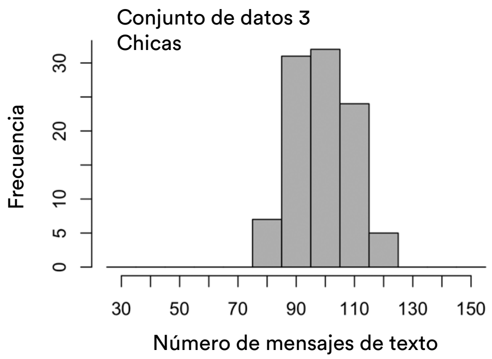
(d) Calcula la puntuación media para cada tipo de chocolatina. ¿Cuál sería la favorita si se determinara con la media? ¿Qué chocolatina sería la menos favorita con este método?

(e) Dibuja un gráfico de barras de la distribución de puntuaciones para cada chocolatina. ¿Cuál crees que sería la favorita si usas este gráfico para compararlas?

Ejemplo 4: Variabilidad

(de la tarea «Describing Distributions» de la web de *Illustrative Mathematics*)

El conjunto 3 contiene datos del número de mensajes de texto enviados en un mes por 100 chicas que tienen teléfono móvil. El conjunto 4 contiene datos del número de mensajes enviados en un mes por 100 chicos que tienen teléfono móvil. En la figura 70 se presentan los histogramas de ambos conjuntos de datos (el alumnado avanzado del nivel A puede responder estas preguntas).



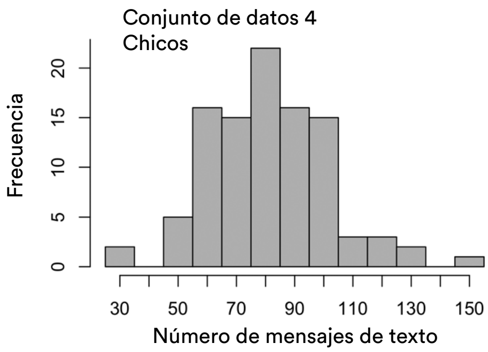


Figura 70. Histogramas para chicas y chicos

(a) Describe la distribución de datos del número de mensajes de texto para las chicas (conjunto de datos 3). Asegúrate de hacer comentarios sobre el centro, la dispersión y la forma general.

(b) ¿Los conjuntos de datos 3 y 4 están centrados aproximadamente en el mismo sitio? Si no es así, ¿cuál tiene el centro mayor?

(c) ¿Qué conjunto de datos está más disperso, el 3 o el 4?

(d) Como promedio, ¿quién envió más mensajes de texto, las chicas (conjunto 3) o los chicos (conjunto 4)?

Ejemplos de evaluación del nivel B

Ejemplo 1: Aleatoriedad

(del proyecto LOCUS)

El alumnado quería investigar si la distancia que puede saltar un estudiante (masculino) se ve afectada por el hecho de tener un objetivo hacia el que dar el salto. Decidió realizar un experimento comparando dos grupos. Un grupo está formado por estudiantes que saltarán hacia un objetivo fijo; y el otro grupo, por estudiantes que saltarán sin un objetivo fijo. Hay 28 estudiantes disponibles para el experimento.

En unas pocas frases, describe cómo dividirías aleatoriamente a los 28 estudiantes para formar los dos grupos.

La tarea puede ampliarse proporcionando los diagramas de caja que muestren los resultados del experimento (véase la figura 71).

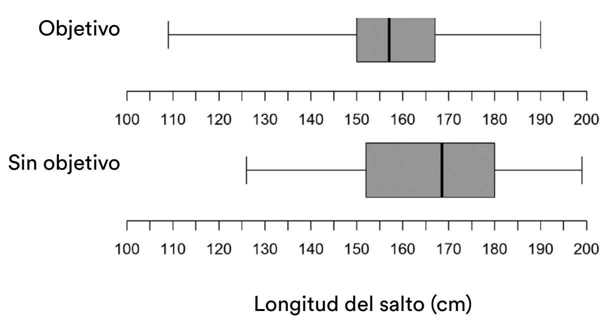


Figura 71. Diagramas de caja para los saltos de los estudiantes, con y sin objetivo

Se puede pedir al alumnado lo siguiente:

«Escribe una conclusión sobre si las distancias que los estudiantes saltaron se vieron afectadas por el hecho de tener un objetivo. Justifica tu conclusión».

Ejemplo 2: Representaciones adecuadas de los datos y comprobación de conjeturas

La tabla 29 muestra los datos que se recogieron de un grupo de 4.º curso de educación primaria.

Tabla 29. Datos recogidos de un grupo de 4.º curso

Nombre	Edad	Número de mascotas	Altura (pulgadas)	Peso (libras)[39]
Mario	9	2	50	68
Lisa	10	0	54	77
Kiko	9	1	52	73
Juan	11	0	57	83
Josef	10	4	52	71
Beatriz	10	2	55	78
Carlos	9	3	51	71
Jeannie	11	1	58	85
David	9	0	52	70
Lynn	10	3	53	75

Haz una conjetura sobre dos características del alumnado que podrían estar relacionadas. Realiza un gráfico o una tabla para comprobar tu conjetura. Explica si los datos respaldan o no tu conjetura.

Ejemplo 3: Comparación de dos conjuntos de datos cuantitativos continuos y obtención de conclusiones

(del proyecto LOCUS)

El año pasado, la ciudad de Gainesville organizó dos carreras el día de Año Nuevo. Los participantes eligieron entre una carrera de 5 km (3,1 millas) o una media maratón (13,1 millas): 134 personas corrieron los 5 km y 224 personas corrieron la media maratón. El tiempo por milla, que es el promedio de tiempo que le lleva a una persona correr una milla, se calculó dividiendo el tiempo que le tomó a cada persona completar la carrera entre la distancia de la carrera. Los histogramas de la figura 72 muestran las distribuciones de los tiempos por milla (en minutos por milla) para las personas que corrieron las dos carreras.

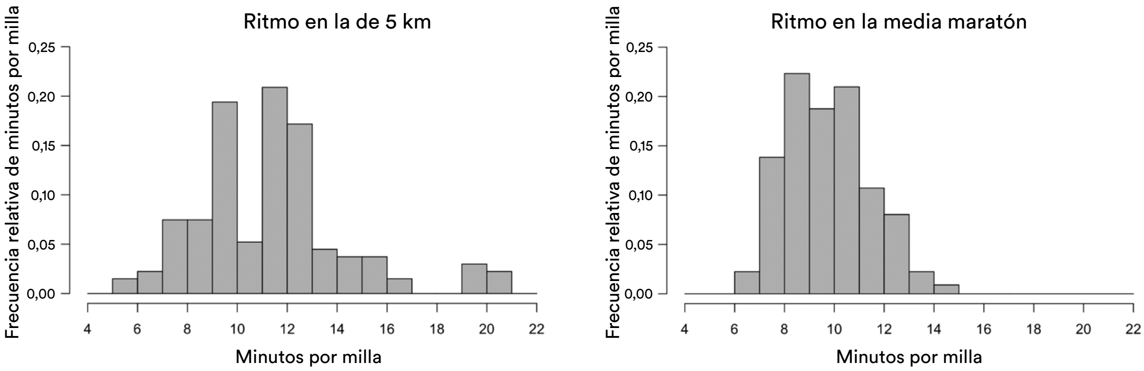


Figura 72. Histogramas de quienes corrieron

[39] N. de T.: 1 libra (lb) equivale a 453,59 g, aproximadamente.

(a) Jaron predijo que los tiempos por milla de quienes corrieron los 5 km serían más uniformes que los tiempos por milla de quienes corrieron la media maratón. ¿Apoyan los datos la afirmación de Jaron? Explica por qué sí o por qué no.

(b) Sierra predijo que, como promedio, el tiempo por milla de quienes corrieron la media maratón sería mayor que el tiempo por milla de quienes corrieron los 5 km. ¿Apoyan los datos la afirmación de Sierra? Explica por qué sí o por qué no.

(c) Recuerda que los participantes eligieron individualmente correr en una de las dos carreras. Según estos datos, ¿es razonable concluir que el tiempo por milla de una persona sería menor cuando corre una media maratón que cuando corre los 5 km? Explica por qué sí o por qué no.

Ejemplo 4: Pasando por el proceso de resolución de problemas estadísticos

(del proyecto LOCUS)

A los miembros del consejo de estudiantes de una gran escuela secundaria se les ha pedido que recomienden una actividad para incluir en las clases de Educación Física el próximo año. Deciden encuestar a 100 estudiantes y preguntarles cuál de las siguientes actividades prefieren: *kickball*, tenis, yoga o danza.

(a) ¿Qué pregunta debería utilizarse en la encuesta? Redacta la pregunta tal como aparecería en la encuesta.

(b) Describe el proceso que usarías para seleccionar una muestra de 100 estudiantes que respondan tu pregunta.

(c) Crea una tabla o gráfico que resuma las respuestas posibles de la encuesta. La tabla o el gráfico deben ser adecuados para esta situación.

(d) ¿Qué actividad debería recomendar el consejo de estudiantes para incluir en las clases de Educación Física el próximo año? Justifica la elección basándote en tu respuesta a la parte (c).

Ejemplo 5: La media como punto de equilibrio y la DAM

(del proyecto LOCUS)

Dos equipos de fútbol se enfrentarán en un partido del campeonato de la ciudad. Durante la temporada, cada equipo ha jugado 10 partidos, con un promedio de 3 goles por partido. Los dos gráficos de puntos de la figura 73 muestran el número de goles anotados por cada equipo en cada partido durante la temporada.

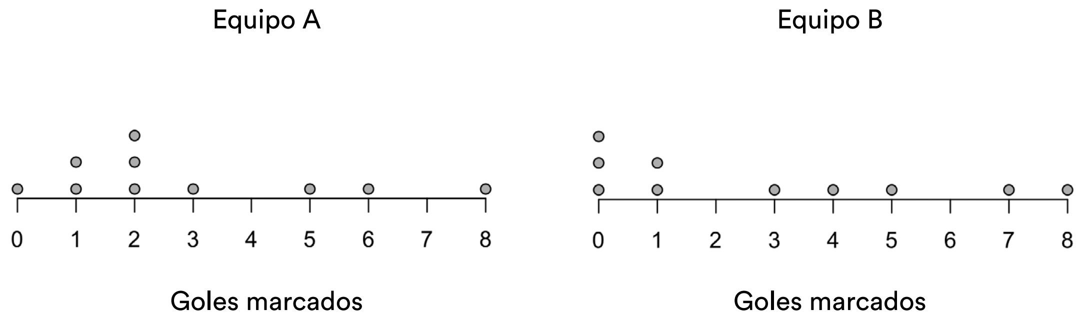


Figura 73. Gráficos de puntos para los goles marcados

(a) Sara, sumando todos los goles marcados por el equipo A y dividiendo entre 10, halló que la media del equipo fue de 3 goles por partido. Mediante los datos del gráfico de puntos muestra que 3 goles es el punto de equilibrio para los goles marcados por el equipo A.

(b) La DAM (desviación absoluta media) del equipo A es de 2 goles. ¿Qué nos dice la DAM sobre la variabilidad del número de goles marcados por el equipo A?

(c) A partir de los gráficos de puntos, ¿qué equipo ha mostrado mayor variabilidad en el número de goles marcados por partido en la temporada? Explícalo.

Ejemplos de evaluación del nivel C

Ejemplo 1: Interpretar un intervalo y el efecto del tamaño muestral en el margen de error

(del proyecto LOCUS)

Lindsey quiere estimar, mediante un intervalo de confianza, la diferencia entre la proporción de mujeres y de hombres de su instituto que han cursado alguna materia en un nivel avanzado.[40] Para ello, elige al azar a 50 chicas y a 50 chicos de su instituto, y les pregunta individualmente si han cursado o no alguna de estas materias. De las 50 chicas, 23 responden que sí; y de los 50 chicos, 19 responden que sí.

El intervalo de confianza que se obtiene al 95 % para la diferencia entre las proporciones de mujeres y de hombres de su escuela que han cursado una materia avanzada es 0,08 ± 0,19.

(a) Interpreta el intervalo de confianza en el contexto de este estudio.

(b) La directora de la escuela de Lindsey está interesada en los resultados, pero le sugiere que aumente los tamaños de las muestras a 100 mujeres y 100 hombres. ¿Qué efecto tendrá incrementar los tamaños muestrales en el intervalo de confianza de Lindsey?

Ejemplo 2: Obtener conclusiones sobre la relación entre dos variables cualitativas

(del proyecto LOCUS)

El Departamento de Salud quiere investigar si existe alguna asociación entre comer en restaurantes de comida rápida y el género. Realizan una encuesta a 100 personas seleccionadas al azar y les hace la siguiente pregunta: «¿Comes en un restaurante de comida rápida al menos una vez por semana?».

(a) ¿Qué tipo de datos (cualitativos o numéricos) se obtendrán con esta pregunta?

De las personas que respondieron, 60 eran hombres. El 54 % de las 100 personas encuestadas indicó que come en un restaurante de comida rápida al menos una vez por semana.

(b) Si no existiera asociación entre el género y el hecho de comer en restaurantes de comida rápida, ¿qué porcentaje de hombres sería previsible que comiera en un restaurante de este tipo al menos una vez por semana? Explica tu razonamiento.

[40] N. de T.: El original dice: «who have taken an honors class». Este tipo de cursos (*honors class*) cubren materias que están diseñadas para estudiantes que han demostrado un rendimiento académico excepcional. Para estas materias, opcionales, el nivel de exigencia es mayor. Por ello, tiene sentido determinar cuál es la proporción de estudiantes que las han cursado, en el contexto del informe GAISE II.

Los resultados de la encuesta se muestran en la tabla 30. Las personas encuestadas se clasifican por género (hombre o mujer) y por si comen o no en un restaurante de comida rápida al menos una vez por semana.

Tabla 30. Respuestas a la encuesta por género

Come en un restaurante de comida rápida al menos una vez por semana			
Género	**Sí**	**No**	**Total**
Hombre	40	20	60
Mujer	14	26	40
Total	54	46	100

(c) Usa la información de la tabla para responder las siguientes preguntas:

(i) Entre los hombres encuestados, ¿qué porcentaje dijo que come en un restaurante de comida rápida al menos una vez por semana?

(ii) Entre las mujeres encuestadas, ¿qué porcentaje dijo que come en un restaurante de comida rápida al menos una vez por semana?

(d) ¿Parece que haya alguna asociación entre el género y el hecho de comer en un restaurante de comida rápida al menos una vez por semana? Justifica tu respuesta.

Ejemplo 3: Describir la relación entre dos variables cuantitativas mediante la interpretación de la recta de regresión de mínimos cuadrados

(del proyecto LOCUS)

Se midió la estatura (en centímetros) y la envergadura de brazos (también en centímetros) de 31 estudiantes. La relación entre *x* (estatura) e *y* (envergadura) se muestra en el diagrama de dispersión (véase la figura 74). También se proporciona la ecuación de la recta de regresión de mínimos cuadrados correspondiente a esta relación.

$$\text{envergadura estimada} = 4{,}5 + 0{,}977 \text{ altura}$$

(a) Si Mike mide 5 cm más que George, ¿cuál es la diferencia esperada entre las envergaduras de sus brazos? Explica cómo lo haces.

(b) Jane mide 158 cm y tiene una envergadura de brazos de 154 cm. Rhonda mide 163 cm y tiene una envergadura de 165 cm. ¿La recta de regresión de mínimos cuadrados proporciona un valor predicho más preciso para Jane o para Rhonda? Explica tu respuesta.

(c) Doug mide 210 cm de altura. ¿Usarías esta recta de regresión de mínimos cuadrados para predecir su envergadura? Explica tu respuesta.

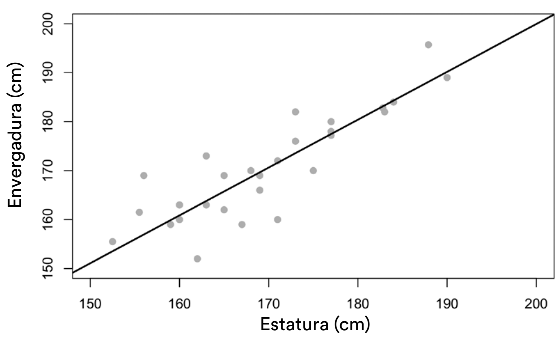


Figura 74. Diagrama de dispersión con la recta de regresión de mínimos cuadrados

Ejemplo 4: Simulación y decisión sobre si el valor observado de un estadístico es inusual o posible

(del proyecto LOCUS)

Stella leyó el siguiente titular en un periódico nacional: «El 30 % del alumnado de bachillerato está a favor de ampliar la jornada escolar». Se preguntó si el porcentaje de estudiantes de su centro que está a favor de ampliar la jornada escolar sería menor del 30 %. Para investigarlo, seleccionó una muestra aleatoria de 50 estudiantes de entre los 1200 de su centro y les preguntó si estaban a favor de ampliar la jornada escolar.

Solo 12 de los estudiantes de la muestra dijeron estar a favor, lo que representa un porcentaje de (12/50) 100 = 24 %. Stella cree que menos del 30 % del alumnado de su centro está a favor. Sin embargo, se pregunta si obtener un 24 % o menos en una muestra sería algo sorprendente si en realidad el porcentaje en la escuela fuera del 30 %.

(a) Para ver qué valores del porcentaje muestral serían esperables si el porcentaje real en la escuela fuera del 30 %, decide usar 1200 bolitas para representar al alumnado de su centro. Usará una bolita roja para representar a cada estudiante que está a favor de ampliar la jornada escolar, y una bolita blanca para cada estudiante que no lo está. ¿Cuántas bolitas rojas y cuántas bolitas blancas debe usar Stella?

Stella colocó todas las bolitas en una caja. Tras mezclarlas, extrajo 50 al azar y calculó el porcentaje de bolitas rojas. Tras devolverlas a la caja, repitió el proceso 99 veces más, de modo que obtuvo un total de 100 porcentajes muestrales. Finalmente, realizó un gráfico de puntos con los resultados (véase la figura 75).

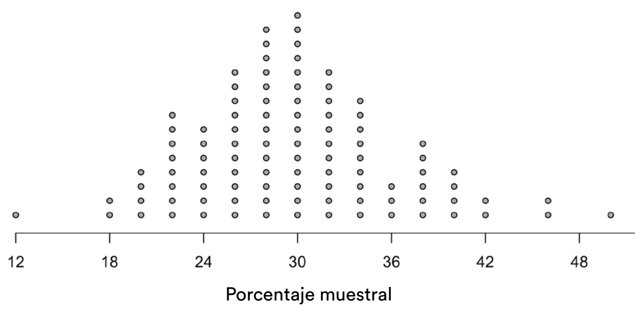


Figura 75. Gráfico de puntos de los porcentajes de bolitas rojas

(b) Si el porcentaje real del centro fuera del 30 %, ¿hasta qué punto sería sorprendente obtener un porcentaje muestral del 24 % o menos? Justifica tu respuesta a partir del gráfico de puntos.

(c) Basándose en los datos de su muestra, ¿debería concluir Stella que el porcentaje de alumnado de su centro que está a favor de ampliar la jornada escolar es menor del 30 %? Explica por qué sí o por qué no.

Referencias

American Statistical Association. (s.f.-a). *Census at School*. https://ww2.amstat.org/censusatschool/

American Statistical Association. (s.f.-b). *Education*. https://www.amstat.org/education

American Statistical Association. (s.f.-c). *STatistics Education Web (STEW)*. https://www.amstat.org/education/stew/statistics-education-web-(stew)

Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016, 23 de mayo). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arnold, P., y Franklin, C. (2021). What makes a good statistical question? *Journal of Statistics and Data Science Education, 29*(1), 122-130. https://doi.org/10.1080/26939169.2021.1877582

Bush, S. B., Roy, G. J., y Jackson, C. (2020). *Catalyzing change in middle school Mathematics: Initiating critical conversations. National Council of Teachers of Mathematics.*

Cobb, G. W., y Moore, D. S. (1997). Mathematics, Statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801-823. https://doi.org/10.1080/00029890.1997.11990723

Dominus, S. (2017, 18 de octubre). When the revolution came for Amy Cuddy. *The New York Times Magazine*.

Dubner, S. (Amfitrión). (2019, 2 de octubre). America's Math Curriculum Doesn't Add Up (No. 391) [Episodio de podcast de audio]. En *Freakonomics Radio*. Freakonomics https://freakonomics.com/podcast/americas-math-curriculum-doesnt-add-up-ep-391/

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R., y Spangler, D. A. (2015). *Statistical education of teachers*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/EDU-SET.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., y Scheaffer, R. (2007). *Guidelines for assessment and instruction in Statistics education (GAISE) report: A Pre-K–12 curriculum framework*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEPreK–12_Full.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., y Scheaffer, R. (2018). *Lineamientos para la evaluación y enseñanza en educación estadística, reporte (GAISE). Un marco para el currículo de Pre-K–12* (J. Zapata, Trad.). American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/Spanish.pdf (Obra original publicada en 2007)

GAISE College Report ASA Revision Committee. (2016). *Guidelines for assessment and instruction in statistics education (GAISE). College report 2016*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf

Gelman, A., y Nolan, D. (2002). You can load a die, but you can't bias a coin. *The American Statistician*, *56*(4), 308-311. https://doi.org/10.1198/000313002605

Graham, K., Burrill, G., y Curtis, J. (2018). *Catalyzing change in high school Mathematics: Initiating critical*. National Council of Teachers of Mathematics.

Grant, P. R., y Grant, B. R. (2013). *Data from: 40 years of evolution. Darwin's finches on Daphne Major Island* [Conjunto de datos]. Dryad, https://doi.org/10.5061/dryad.g6g3h

Häusler, N., Haba-Rubio, J., Heinzer, R., y Marques-Vidal, P. (2019). Association of napping with incident cardiovascular events in a prospective cohort study. *Heart*, *105*(23), 1793-1798. https://doi.org/10.1136/heartjnl-2019-314999

Hopfensperger, P., Brown, S., y Kranendonk, H. (2020). *Focus on statistics: Investigations for the integration of Statistics into grades 9-12 Mathematics classrooms*. American Statistical Association.

Huinker, D., Yeh, C., y Marshall, A. M. (2020). *Catalyzing change in early childhood and elementary Mathematics: Initiating critical conversations*. National Council of Teachers of Mathematics.

Kader, G. D., y Franklin, C. A. (2008). The evolution of Pearson's correlation coefficient. *Mathematics Teacher*, *102*(4), 292-299.

Kuiper, S. R. (2010). Incorporating research experience into an introductory college-level Statistics Course. En Reading, C. (Ed.), *Data and context in Statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010)*, Liubliana, Eslovenia. International Statistical Institute. https://icots.info/documents/papers/icots8/ICOTS8_4G1_KUIPER.pdf?1402524970

Leavy, A., Meletiou-Mavrotheris, M., y Paparistodemou, E. (2018). *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking*. Springer. https://doi.org/10.1007/978-981-13-1044-7

Lindgren, M. (s.f.). *Detailed income calculations for Dollar Street*. https://drive.google.com/drive/folders/0B9jWD65HiLUnRm5ZNWlMSU5GNEU

Markow, W., Hughes, D., y Bundy, A. (2018). *The new foundational skills of the digital economy: Developing the professionals of the future*. Burning Glass Technologies; Business-Higher Education Forum. https://www.bhef.com/sites/default/files/BHEF_2018_New_Foundational_Skills.pdf

National Council of Teachers of Mathematics. (s.f.). *More 4 U*. https://www.nctm.org/store/more4u/

National Research Council. (2013). *Next generation science standards: For states, by states*. The National Academies Press. https://doi.org/10.17226/18290

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, *506*(7487), 150-152. https://doi.org/10.1038/506150a

Peck, R., Gould, R., y Miller, S. (2013). *Developing Essential Understanding of Statistics for Teaching Mathematics in Grades 9-12*. National Council of Teachers of Mathematics.

Pfannkuch, M., y Budgett, S. (2017). Reasoning from an eikosogram: An exploratory study. *International Journal of Research in Undergraduate Mathematics Education*, *3*(2), 283-310.

Rosling Ronnlund, A. (s.f.). *Dollar Street—Photos as data to kill country stereotypes*. Dollar Street. https://www.gapminder.org/dollar-street/about

Rosling Rönnlund, A. (2017, abril). *See how the rest of the world lives, organized by income* [Vídeo].

TED Conferences. https://www.ted.com/talks/anna_rosling_ronnlund_see_how_the_rest_of_the_world_lives_organized_by_income

Silverman, M. E., Murray, T. J., y Bryan, C. S. (2008). *The quotable Osler*. American College of Physicians.

Singer, N., y Metz, C. (2019, 19 de diciembre). Many facial-recognition systems are biased, Says

U.S. study. *The New York Times*. https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html

The Learning Network. (2018, 9 de enero). What's Going On in This Graph.*The New York Times*. https://www.nytimes.com/2018/01/04/learning/whats-going-on-in-this-graph-jan-9-2018.html

The Learning Network. (2019, 25 de septiembre). What's Going On in This Graph. *The New York Times*. https://www.nytimes.com/2019/09/19/learning/whats-going-on-in-this-graph-sept-25-2019.html

Wasserstein, R. L., y Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108