

¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

Luis J. Rodríguez-Muñiz (Universidad de Oviedo. España)

Laura Muñiz-Rodríguez (Universidad de Oviedo. España)

Claudia Vásquez (Pontificia Universidad Católica de Chile. Chile)

Ángel Alsina (Universidad de Girona. España)

Fecha de recepción: 29 de mayo de 2020

Fecha de aceptación: 30 de junio de 2020

Resumen

En este artículo se presenta una propuesta de actividades de aula para Educación Secundaria (12-16 años) a partir de la COVID-19, con base en una fundamentación teórica sobre la alfabetización estadística y de datos. En la primera parte, se describen los conocimientos matemáticos y didácticos que debe movilizar el profesorado de matemáticas de esta etapa educativa para promover la alfabetización estadística y de datos a partir de contextos cotidianos y de interés social, junto con herramientas como la manipulación tecnológica para manejar bases de datos de gran tamaño; en la segunda parte, se presentan cuatro experiencias de aula contextualizadas en la COVID-19 para desarrollar la alfabetización estadística y de datos del alumnado. Se concluye que la alfabetización estadística y de datos, junto con la alfabetización probabilística, permiten desarrollar una mirada crítica contra datos manipulados, noticias falsas, etc., y tomar decisiones informadas.

Palabras clave

Alfabetización estadística, alfabetización de datos, conocimiento matemático para la enseñanza, enseñanza de las matemáticas en contexto, COVID-19, Educación Secundaria.

Title

How to promote statistical and data literacy in context? Strategies and resources from COVID-19 for Secondary Education

Abstract

In this article, we present a proposal of classroom activities for Secondary Education (12-16 years old) from COVID-19, based on a theoretical framework on statistical and data literacy. In the first part, we describe the mathematical and didactic knowledge that must be mobilized by the mathematics teachers of this educational stage to promote statistical and data literacy from everyday contexts and of social interest, together with tools such as technological manipulation to handle large databases. In the second part, four classroom experiences contextualized in COVID-19 are presented to develop students' statistical and data literacy. It is concluded that statistical and data literacy, together with probabilistic literacy, allow to develop a critical view against manipulated data, fake news, etc., and to make informed decisions.

Keywords

Statistical literacy, data literacy, mathematical knowledge for teaching, mathematics teaching in context, COVID-19, Secondary Education.



1. Introducción

En el artículo con esta misma pregunta en el título, pero dirigido a la etapa de Educación Primaria (Alsina, Vásquez, Muñiz-Rodríguez y Rodríguez-Muñiz, 2020), se ha hecho hincapié en la necesidad de promover la alfabetización estadística (*statistical literacy*) y la alfabetización probabilística (*probabilistic literacy*) del alumnado desde los primeros niveles de escolarización, con el propósito de que progresivamente adquieran los conocimientos necesarios para ser ciudadanos críticos en el análisis de datos y, a la vez, capaces de tomar decisiones en situaciones de incertidumbre. Este planteamiento competencial, lógicamente, debe proseguir en la etapa de Educación Secundaria, considerando además la alfabetización de datos (*data literacy*), fuertemente influenciada por el auge de la ciencia de datos, vinculada a los macrodatos (*big data*) y a la llamada cuarta revolución industrial (Gleason, 2018), a lo que se deben agregar fenómenos insospechados como la crisis generada por la COVID-19, con repercusiones socioeconómicas de gran alcance. En este sentido, la combinación de conceptos y métodos de estadística e informática, pero también de matemáticas y ciencias de la información, puede ayudar a procesar y comprender la gran cantidad de datos sociodemográficos, económicos, sanitarios y de movilidad derivados de la crisis mundial provocada por dicha pandemia.

Otro aspecto relevante al que se ha hecho referencia en el artículo de Alsina et al. (2020) es, lógicamente, la forma de promover la alfabetización estadística y probabilística. En este sentido, de acuerdo con autores como Hahn (2015) y Muñiz-Rodríguez, Rodríguez-Muñiz y Alsina (2020), entre otros, se ha insistido en que es necesario usar datos reales, significativos y motivadores para el alumnado, en lugar de usar datos precocinados (tomados, por ejemplo, del libro de texto) y plantear problemas sobre algo que no es problemático para el alumnado, omitiendo todas las dificultades derivadas de los procesos de recolección o muestreo. Este planteamiento realista se vincula directamente con los enfoques de aprendizaje basados en proyectos (Batanero y Díaz, 2011), que permiten tomar decisiones sobre cuándo, dónde y cómo muestrear y sobre cómo recopilar y organizar los datos.

Con el propósito de ofrecer una continuidad que no suponga una rotura en la transición entre las etapas de Educación Primaria y Secundaria, en esta misma sección de Propuestas de Aula se presentan dos nuevos artículos entrelazados que tratan, respectivamente, la alfabetización estadística y de datos y la alfabetización probabilística en Educación Secundaria (12-16 años). Se trata de dos documentos estrechamente relacionados porque, muy conscientemente, se quiere romper con una visión fragmentada de la estadística y la probabilidad, entendidas como dos disciplinas que pueden abordarse separadamente en el aula. Por el contrario, es imprescindible tener siempre presente el objeto de la probabilidad y su relación con la estadística tanto descriptiva como inferencial al planificar las prácticas de enseñanza de la estadística y la probabilidad.

Con base en estas consideraciones, en la primera parte de este artículo focalizado en la alfabetización estadística y de datos, se presentan los conocimientos que la investigación recomienda que debería movilizar el profesorado de Educación Secundaria para promover esta alfabetización; y en la segunda parte, se presentan diversas experiencias de aula para trabajar estos conocimientos en contexto, a partir de datos de distinta naturaleza obtenidos de la COVID-19.

2. Conocimientos de estadística en Educación Secundaria

Hoy más que nunca es necesario contar con ciudadanos con un pensamiento crítico, capaces de comprender la información estadística que nos bombardea a diario y que afecta, de una u otra manera, a nuestra cotidianeidad. Ejemplos como el uso de perfiles falsos (*bots*) para generar y difundir en las

redes sociales noticias falsas (*fake news*) evidencian la necesidad de vincular el análisis de datos con la información a la que está expuesta la ciudadanía, yendo desde una alfabetización estadística hacia una alfabetización de datos o *data literacy* (Schield, 2004). La importancia de los datos se ha manifestado de manera creciente en los últimos años, con el auge de los macrodatos (*big data*) y sus aplicaciones en todos los ámbitos sociales y económicos. Obviamente, en la Educación Secundaria no vamos a formar analistas de datos, pero sí debemos encaminar al alumnado para poder desenvolverse en la llamada *era de la información* (Castells, 2005) con suficientes herramientas estadísticas, sea cual sea su ocupación en el futuro.

Para acometer esta tarea de alfabetización, es necesario que el profesorado cuente con las herramientas para desarrollar una educación estadística que vaya más allá de los aspectos mecánicos y computacionales de los métodos estadísticos descriptivos, que fomente una comprensión conceptual sobre lo procedimental, inculcando en su alumnado la capacidad de pensar estadísticamente (NCTM, 2013). Así, se espera que para la Educación Secundaria el nivel de formalización de las ideas estadísticas vaya en aumento, con el propósito de ayudar al alumnado a desarrollar su alfabetización estadística para avanzar progresivamente hacia la adquisición del sentido estadístico (Batanero, Díaz, Contreras y Roa, 2013). Además, debemos avanzar hacia la alfabetización de datos, cuya importancia ha sido recientemente reconocida por la OCDE (2019), que la ha señalado como uno de los grandes objetivos de su Agenda 2030. Así, la OCDE pone el énfasis en la obtención de datos fiables y de calidad que, mediante su organización y, particularmente, su visualización produzca información interpretable y significativa.

2.1. Conocimientos acerca de la estadística y su enseñanza

Desde esta perspectiva, la estadística debe ser entendida como una herramienta que contribuye a la resolución de problemas, donde el contexto juega un papel fundamental en su enseñanza, pues “los datos no son sólo números, sino números en un contexto. En matemáticas el contexto oscurece la estructura. En el análisis de datos, el contexto proporciona significado” (Moore y Cobb, 1997, p. 801). Por tanto, es necesario abordar adecuadamente el proceso de resolución de problemas estadísticos, pues “la estadística es inseparable de sus aplicaciones, y su justificación final es su utilidad en la resolución de problemas externos a la propia estadística” (Batanero y Díaz, 2011, p. 21). En este sentido, la enseñanza de la estadística se puede abordar a través de proyectos enfocados en contextos realistas (incluso cuando sean versiones simplificadas de un problema) que lleven al alumnado hacia el aprendizaje de conceptos estadísticos, el empleo de técnicas de cálculo, la mejora de sus capacidades de argumentación, la formulación de conjeturas y la reflexión en torno a dicho contexto. El desarrollo del sentido estadístico no debe ceñirse a procesos algorítmicos, sino que requiere introducir el pensamiento crítico mediante el uso de problemas contextualizados, datos reales y problemas abiertos en los que, mediante el uso imprescindible de la tecnología, el alumnado deba tomar decisiones en ambientes de incertidumbre (Rodríguez-Muñiz, Díaz y Muñiz-Rodríguez, 2019). En definitiva, la metodología de enseñanza de la estadística por medio de proyectos busca que el alumnado aborde la resolución de un problema real y con significado de manera similar a como se hace en la estadística profesional. Desde este punto de vista, en las Directrices para la Evaluación y Enseñanza en Educación Estadística (marco GAISE) definidas por la *American Statistical Association* (GAISE College Report ASA Revision Committee, 2016; Franklin et al., 2005) se considera la resolución de problemas estadísticos como un proceso investigativo que involucra cuatro componentes que se encuentran interrelacionados:

1. *Formulación de una pregunta estadística, es decir, una pregunta que pueda responderse mediante datos.* Es importante diferenciar entre las preguntas que anticipan una respuesta determinista de las que anticipan una respuesta basada en la variabilidad: por ejemplo, la pregunta ¿cuál es mi peso?



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

no es una pregunta estadística, ya que puede ser respondida con un solo dato, mientras que la pregunta ¿cuál es el peso del alumnado de primero de Secundaria? es una pregunta estadística, pues el hecho de encontrarnos con alumnado de diferente peso implica que nos anticipamos a una respuesta basada en un conjunto de datos que varían dependiendo de los individuos analizados. Es importante iniciar el planteamiento de preguntas estadísticas a partir de contextos cercanos al alumnado, que tengan significado para él y que, a la vez, le permita desarrollar el sentido de los datos, como, por ejemplo, preguntas estadísticas sobre el grupo de la clase. Posteriormente, es importante generar preguntas estadísticas más amplias que surjan desde el propio alumnado, comenzando a diferenciar entre una población, un censo, una muestra, tipos de variables, así como comprender las diferencias entre varias clases de estudios estadísticos y sentar las bases del razonamiento inferencial.

2. *Diseño de un plan para recopilar datos útiles, implementación del plan y recopilación de datos, reconociendo la variabilidad en los datos.* En este sentido, es importante plantear actividades que, a partir de un número reducido de datos, permita al alumnado entender la variabilidad natural de individuo a individuo, así como la importancia de diseñar un plan que posibilite recoger los datos adecuados que permitan responder a la pregunta estadística planteada. Seguidamente, se debe transitar hacia el diseño e implementación de encuestas que proporcionen respuestas a preguntas estadísticas más sofisticadas, que involucren el muestreo aleatorio y el no aleatorio, los experimentos comparativos, etc. También se debe alcanzar una comprensión en profundidad de lo que constituye una buena práctica de realizar una encuesta por muestreo, un experimento, y un estudio observacional. Finalmente, es imprescindible experimentar que, en el mismo estudio estadístico, muestras diferentes proporcionan información diferente, de modo que la esencia de la variabilidad se haga evidente. Si, por el contrario, siempre que se haga un estudio estadístico, se trabaja con la información de una sola muestra, la variabilidad quedará reducida a una premisa hipotética.

3. *Análisis de los datos*, a través de la creación y exploración de varias representaciones de la distribución de los datos, con el propósito de identificar y describir patrones que permitan dar cuenta de la variabilidad de los datos y resumir características de la distribución. En lo que respecta al análisis de los datos, se debe fomentar el uso de diversas técnicas y representaciones que permitan:

- Comprender el significado de los datos cuantitativos y cualitativos, de los datos unidimensionales y bidimensionales, y del término variable.
- Seleccionar, crear, utilizar y comprender representaciones gráficas de datos apropiadas, incluyendo las representaciones más elementales, como pictograma, gráficas de puntos (o *dotplot*), gráfico de barras, poligonal de frecuencias o diagrama de sectores; y otras más avanzadas, como histograma, diagrama de dispersión (o nubes de puntos), diagrama de caja y bigotes (o *boxplot*), así como otras gráficas menos presentes en el ámbito escolar pero muy relevantes en la sociedad de la información (cartogramas, coropletas, gráficos de burbujas, nubes de términos, mapas de calor, diagramas de radar, diagramas de área polar, etc.).
- Discutir y comprender la correspondencia entre conjuntos de datos y sus representaciones gráficas, y entre diferentes representaciones gráficas.
- Efectuar cálculos estadísticos básicos y comprender la diferencia entre un estadístico, como variable aleatoria, y un parámetro como valor del estadístico para una muestra concreta.
- Respecto a los datos unidimensionales, representar gráficamente la distribución, describir su forma y seleccionar y elaborar un resumen estadístico, incluyendo las características más notables.
- Reconocer cómo las transformaciones lineales de datos unidimensionales afectan a la forma, la centralización y la dispersión.

- Con relación a los datos bidimensionales, representarlos mediante una nube de puntos o diagrama de dispersión, describir su forma, y determinar los coeficientes y la recta de regresión, y los coeficientes de correlación y determinación, utilizando herramientas tecnológicas.
- Identificar tendencias en datos bidimensionales, comparar las relaciones identificadas en términos de la correlación, distinguir correlación de causalidad y de relación funcional y hallar, en su caso, funciones que modelen estas relaciones.
- Utilizar simulaciones para explorar la variabilidad de muestras estadísticas de una población conocida.

4. *Interpretación de resultados*, proporcionando una respuesta estadística a la pregunta planteada que tenga en cuenta la variabilidad de los datos, permitiendo mirar más allá de los datos. De esta manera, este componente debe focalizarse en los siguientes aspectos:

- Realizar inferencias sobre un grupo pequeño como, por ejemplo, el alumnado de una clase.
- Describir e interpretar diferencias entre dos o más grupos con respecto al centro, la dispersión y la forma de las distribuciones.
- Hallar, utilizar e interpretar medidas de centralización y de dispersión: media, mediana, moda, rango, rango intercuartílico, desviación típica, y varianza.
- Reconocer que una muestra puede o no ser representativa de una población mayor.
- Comprender interpretaciones básicas de las medidas de asociación.
- Utilizar observaciones relativas a las diferencias entre dos o más muestras, para formular conjeturas sobre las poblaciones de las que se han extraído.
- Distinguir entre un estudio observacional y un experimento diseñado.
- Interpretar el tipo de relación entre dos variables, distinguiendo entre asociación y correlación.
- Reconocer la variabilidad muestral en el comportamiento de distintos estadísticos como la media muestral o la proporción muestral.
- Realizar inferencias informales sobre la población basándose en la información obtenida de la muestra y plantear las bases de la inferencia estadística formal, comprendiendo la relación entre población y muestra.

En definitiva, se trata de abordar la enseñanza de la estadística desde estos componentes para desarrollar una comprensión de los elementos que la definen: datos, gráficos, variabilidad, distribución, asociación y correlación, probabilidad, muestreo e inferencia (Burrill y Biehler, 2011).

2.2. Otros conocimientos didácticos

Además de los conocimientos específicos sobre la estadística y su enseñanza, el profesorado de matemáticas de Educación Secundaria debe tener un conocimiento didáctico profundo sobre la didáctica específica de la matemática y, más concretamente, de la didáctica de la estadística y la probabilidad. Este conocimiento requiere tener presente la diferencia esencial entre la matemática determinista y la que maneja incertidumbre. Los problemas, los ejemplos y las actividades son necesariamente diferentes. La importancia de los datos y de la variabilidad la hemos subrayado anteriormente, pero también es muy



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

importante tener presente el objeto de la probabilidad y su relación con la estadística tanto descriptiva como inferencial (Figura 1).

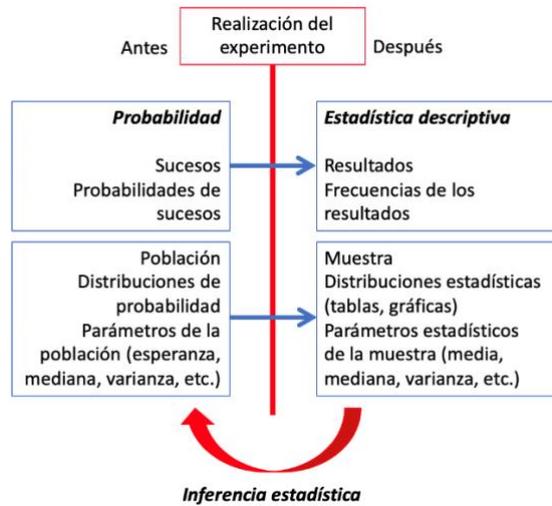


Figura 1. Esquema del tipo de tratamiento matemático de los experimentos aleatorios. Fuente: Elaboración propia.

Aunque la probabilidad se trata con mayor profundidad en el artículo sobre alfabetización probabilística de esta misma sección, en la Figura 1 se establecen puentes entre la estadística y la probabilidad con el propósito de fijar las estrechas relaciones que existen entre ambas materias, tal como se ha indicado en la introducción. Desde este prisma, en ocasiones se escucha la idea, errónea, de que la probabilidad sirve para predecir el resultado de un experimento aleatorio: si fuera así, todos los docentes de probabilidad se habrían hecho ricos con los juegos de azar. La probabilidad nos ayuda a estimar, antes de que se produzca un experimento aleatorio, el grado de creencia (objetiva o subjetiva) que tenemos acerca de que se produzca un determinado resultado del experimento. Pero ese grado, incluso aunque sea muy elevado, no predice su ocurrencia. Después de realizar el experimento, el resultado se habrá producido o no, pero una sola repetición del experimento no respalda ni refuta una probabilidad, sería necesario tener un gran número de repeticiones para poder revisar nuestra probabilidad. Por ejemplo, un equipo de fútbol modesto puede vencer en un partido a un equipo multicampeón, aunque sea muy improbable, podría ocurrir. Ahora bien, salvo que en varios enfrentamientos siga venciendo el equipo modesto, seguiríamos manteniendo que es más probable que venza el campeón.

Por otro lado, la estadística descriptiva tiene lugar después de la realización del experimento, para recopilar, organizar y analizar los datos del resultado de ese experimento cuantas veces haya sido realizado. Y, finalmente, la inferencia (formal o informal) nos permite dar un giro a la situación, utilizando los datos de lo que ha ocurrido en un experimento para emitir, y en su caso contrastar, estimaciones, conjeturas e hipótesis acerca del comportamiento del experimento en un conjunto mayor que el estudiado (en la población) o ante una posible nueva ejecución del mismo. Por ello, los problemas que planteemos en estadística y en probabilidad, deben estar encaminados no al cálculo de parámetros o valores, sino a la argumentación y a la toma de decisiones en ambientes de incertidumbre de contextos próximos al alumnado, que le permitan manejarse con soltura respecto a qué información maneja en cada caso, antes o después de realizar un experimento aleatorio, y a cómo los resultados de experimentos

anteriores y el estudio de datos muestrales pueden contribuir a inferir resultados de realizaciones experimentales futuras y de datos poblacionales.

En el ámbito de la estadística y la probabilidad en Educación Secundaria tienen una importancia capital los conocimientos tecnológico-didácticos del profesorado. Como señalan Batanero y Díaz (2011) en esta etapa la enseñanza de la estadística ha de pasar, necesariamente, por el uso de bases de datos públicas y de tecnología informática para la organización y el análisis de los datos estadísticos y para su visualización (especialmente en el caso de gráficas estadísticas no convencionales, como las mencionadas anteriormente, y, sobre todo, en el caso de grandes cantidades de datos). Además, las simulaciones son fundamentales para explorar y comprender el significado de la probabilidad, contrastando los resultados teóricos con los resultados obtenidos de la simulación. No es necesario un complicado software específicamente estadístico, sino que con hojas de cálculo se pueden acometer simulaciones y representaciones gráficas novedosas.

También hay múltiples recursos *online* que favorecen no sólo la simulación, el cálculo y la representación, sino también la interacción con los productos obtenidos. Por ejemplo, podemos encontrar representaciones gráficas en las que apreciamos cómo al aumentar o reducir el tamaño de una barra se modifica la media, o cómo al mover varios de los puntos del diagrama de dispersión cambia la correlación entre los datos. Esta manipulación tecnológica es imprescindible por dos motivos: porque incorpora al ámbito digital las ventajas didácticas de la manipulación y porque permite acercarse al manejo de grandes datos sin necesidad de realizar cálculos o representaciones manuales.

Como señalábamos al principio, no es el objetivo convertir al alumnado de Secundaria en científicos de datos, pero manejando bases de gran tamaño, lo que se denomina *little big data* en el ámbito educativo (Marchiori, 2016), conseguiremos sentar las bases para su alfabetización de datos (*data literacy*) y, al tiempo, daremos sentido a la utilización de la estadística en contextos cotidianos y de interés social.

3. Propuestas de actividades de aula a partir de la COVID-19, organizadas por edades.

A continuación, se describe un conjunto de propuestas de actividades contextualizadas en la COVID-19, cuyo objetivo es promover la alfabetización estadística y de datos en pos del desarrollo del sentido estadístico en Educación Secundaria. Concretamente, se muestran cuatro experiencias, organizadas progresivamente por edades desde los 12 a los 16 años, que se corresponden a los cursos de Educación Secundaria en la mayor parte de los países. La estructura que se sigue es siempre la misma: en primer lugar, se indica el nivel o edad de implementación; en segundo lugar, los contenidos vinculados a la alfabetización estadística y de datos que se abordan en cada experiencia; y, finalmente, una descripción de cada actividad con diversas orientaciones para su implementación en el aula.

Experiencia 1: La COVID-19 en dos gráficas no convencionales

Nivel: 12-13 años.

Contenidos implícitos: Tipos de variables. Gráficos estadísticos. Medidas de centralización (media y moda).

Descripción de la actividad: Se presenta una primera gráfica de burbujas, con los datos de número total de positivos, detectados a través de pruebas PCR (*Polymerase Chain Reaction*) en las distintas



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

Comunidades Autónomas (CCAA) de España (Figura 2). La gráfica de burbujas, o *bubble chart*, representa los valores de una variable asociando una burbuja o círculo de tamaño proporcional a la frecuencia de cada valor. Se plantean inicialmente al alumnado preguntas de reconocimiento: ¿qué información se representa en el gráfico? ¿Por qué unas burbujas son mayores que otras? Es muy importante dejar espacios para expresar la opinión y discutirla sin obsesionarse en validar las respuestas del alumnado. Ha de vincularse la respuesta a dos nociones fundamentales: variabilidad (observando que el número de positivos varía de unas CCAA a otras) y recuento (contamos el número de positivos por comunidad, siendo este el valor que genera la burbuja). También podemos preguntar por alguna medida que obligue a relacionar distintos casos y no sólo a identificar valores: ¿qué comunidad autónoma tiene el mayor número de casos? ¿Cuántos casos son? Con estas dos preguntas abordamos algunos de los errores más habituales en la identificación de la moda: confundir el valor más repetido con su frecuencia e identificar la moda con la mayor frecuencia y no con el valor más frecuente.

Posteriormente podemos plantearnos las conexiones entre diferentes tipos de gráficas: ¿de qué otra forma podemos representar esta información? Dependiendo de la habilidad de la clase, se puede o bien construir con el alumnado una gráfica de barras o bien presentársela ya hecha. En algún momento ha de surgir, espontáneamente en el alumnado o motivado por el profesorado, cómo ordenar las CCAA en el eje de abscisas para representar el diagrama de barras, de donde podemos inferir la necesidad de adoptar un código o un convenio que permita utilizar la representación de barras (orden alfabético, de mayor a menor número de casos positivos, etc.). Al no existir un convenio universalmente admitido, es interesante discutir con el alumnado cómo aprecian mejor las diferencias entre CCAA. Otra alternativa es la representación con un gráfico circular (de sectores o de coronas circulares), para lo cual sería necesario calcular el total de positivos en España y qué porcentaje o proporción sobre el total representa el número de positivos en cada una de las CCAA. En este caso, suele estar más generalizada la ordenación de los sectores de mayor a menor proporción, pero deberíamos utilizar el ejemplo para dar libertad a la argumentación del alumnado respecto a qué orden seguir en la representación.

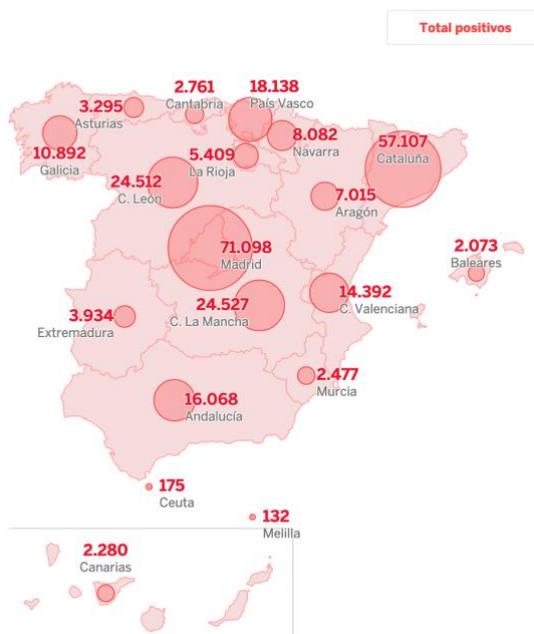


Figura 2. Gráfico del número de positivos por comunidad autónoma a 17 de mayo de 2020. Fuente: www.elpais.es.

Finalizamos proponiendo calcular una media a partir de los datos de la Figura 2: sumando todos los casos y dividiendo entre el número de CCAA, ¿qué significado tendría ese valor obtenido? Se trataría del número medio de casos por comunidad autónoma, pero habiéndolo calculado como una media aritmética simple, ¿tendría algún significado respecto al total nacional si no hemos tenido en cuenta la población de cada comunidad? Es obvio que la respuesta es no, ya que necesitaríamos ponderar los valores de las CCAA por sus poblaciones. Ilustrando este ejemplo, contribuimos a mirar críticamente la información gráfica y a valorar que se pueden calcular diferentes valores medios pero que no siempre tienen una traducción fácil o interpretable en el contexto.

Experiencia 2: Incidencia de la COVID-19 por CCAA

Nivel: 13-14 años.

Contenidos implícitos: Tipos de variables. Gráficos estadísticos. Medidas de centralización (media y moda).

Descripción de la actividad: Iniciamos la actividad presentado una coropleta (Figura 3), mapa que asocia un grado de intensidad de un determinado color en función del valor observado en cada región geográfica de una variable normalmente cuantitativa continua. También usaremos, a modo de comparación, la anterior Figura 2. En esta ocasión planteamos la identificación de la variable y su comprensión a partir de la leyenda: IA 14d casos/100 000 hab. Explicamos las abreviaturas: IA 14d quiere decir incidencia acumulada en los últimos 14 días, es decir el número de casos identificados durante ese periodo, por comunidad autónoma, y medido de forma relativa a la población total por cada 100 000 habitantes.

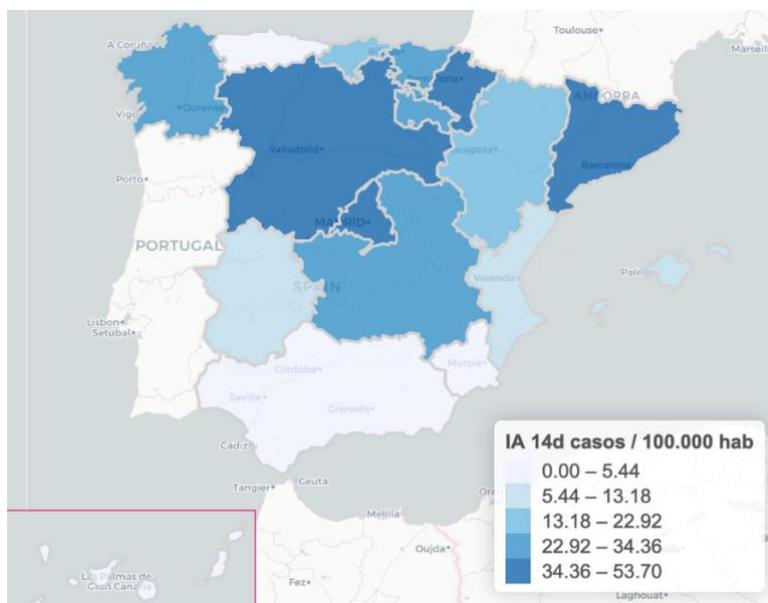


Figura 3. Incidencia Acumulada (IA) de la COVID-19 por CCAA a 17 de mayo de 2020. Fuente: Instituto de Salud Carlos III.

Planteamos cuestiones para profundizar en la comparación y el análisis de gráficos: ¿en qué se parece el gráfico de la Figura 3 al de la Figura 2? y ¿en qué se diferencia? Podríamos comenzar señalando

¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

que en el primer caso la variable está identificada como discreta, y en el segundo, al ser rangos en una escala de cinco colores, está tratada como variable continua, agrupada por intervalos. Además, el ejemplo de Andalucía, con más de 16 000 casos sobre una población superior a 8 millones de habitantes, nos sirve para identificar las diferencias entre cantidades totales (Figura 2) y relativas (Figura 3), y recíprocamente, en el caso de la Comunidad de Navarra, donde la incidencia en términos absolutos es alta, se observa con mayor nitidez si se establece la comparación relativa al tener una población pequeña. Sobre la Figura 3 podemos identificar la moda: ¿qué comunidad tuvo una mayor incidencia acumulada en ese periodo? Obviamente, la gráfica nos muestra una distribución multimodal, lo que podemos aprovechar para recordar la no unicidad de la moda como medida de centralización. A partir de la escala, podemos plantear la pregunta: ¿se puede calcular la incidencia acumulada (IA) media por Comunidad Autónoma (CCAA)?, guiando así el proceso de identificar y calcular los puntos medios de cada intervalo de la escala cromática, y multiplicarlo por el número de CCAA que están en cada intervalo de color.

IA	Marca de clase (x_i^*)	Número observado de CCAA (n_i)
0-5.44	2.72	4
5.44-13.18	9.31	3
13.18-22.92	18.05	2
22.92-34.36	28.64	4
34.36-53.70	58.35	4
		$N=17$

Tabla 1. Distribución de la Incidencia Acumulada (IA) de la COVID-19 por CCAA a 17 de mayo de 2020. Fuente: Elaboración propia a partir de datos del Instituto de Salud Carlos III.

Con los datos del ejemplo, reproducimos en la Tabla 1 el cálculo de las marcas de clase o puntos medios de cada intervalo y el conteo respecto al número de CCAA observadas en cada intervalo, lo que proporciona una IA media de 24.87 casos por cada 100 000 habitantes:

$$\overline{IA} = \frac{\sum x_i^* n_i}{N} = \frac{2.72 \cdot 4 + 9.31 \cdot 3 + 18.05 \cdot 2 + 28.64 \cdot 4 + 58.35 \cdot 4}{17} = 24.87$$

Es importante señalar que el valor obtenido es la incidencia acumulada media por CCAA y no la incidencia media a nivel nacional, ya que para calcularla sería necesario conocer la población de cada comunidad y obtener el valor como una media ponderada por dicha población.

También podemos utilizar este procedimiento para recordar que la sustitución de todo el intervalo por su punto medio o marca de clase es un procedimiento que sintetiza la información que pueda haber en el intervalo, simplificando el cálculo, pero que, evidentemente, pierde información. Por ejemplo: vemos que, en la sección más oscura de la escala, de 34.36 a 53.70 casos por 100 000 hab., hay cuatro CCAA, pero realmente ignoramos, con estos datos, cuál es la incidencia real en cada una de ellas, por eso, a los efectos de hacer la media, sólo podemos simplificar tomando cada sección de la escala cromática por su punto medio.

Asimismo, también podemos utilizar este ejemplo para comentar que los distintos programas de representación habitualmente realizan de manera automática la distribución en intervalos de la variable; así, por lo general, las distribuciones suelen hacerse en intervalos de la misma amplitud, salvo que haya

grandes diferencias entre los valores de la variable, en cuyo caso se aconseja utilizar intervalos diferentes para evitar que haya muchos intervalos sin observaciones. Este es probablemente el motivo por el que la escala no está formada por intervalos de igual amplitud.

Finalizamos haciendo de nuevo una comparación con la Figura 2, en ella también podríamos haber calculado una media: sumando todos los casos y dividiendo entre el número de CCAA, pero ¿qué significaría esa media? ¿Tendría algún significado si al calcularla no hemos tenido en cuenta la población de cada CCAA? Si queremos saber dónde es mayor la incidencia de la COVID-19, ¿qué media nos resulta más útil, la obtenida con la Figura 2 o con la Figura 3? ¿Por qué? Así, contribuimos a mirar críticamente la información gráfica y a considerar que se pueden calcular muchos valores medios, pero que, como ya indicamos en la experiencia 1, no siempre tienen una traducción fácil o interpretable en el contexto.

Experiencia 3: ¿La enfermedad ataca por igual a hombres y a mujeres?

Nivel: 14-15 años.

Contenidos implícitos: Histogramas (pirámides de población), medidas de centralización y dispersión, representatividad de la media, comparación de distribuciones, datos agrupados.

Descripción de la actividad: Comenzamos la actividad accediendo al portal dedicado a la COVID-19 del Instituto de Salud Carlos III (<https://cnecovid.isciii.es/covid19/>), con el objeto de desarrollar así la competencia tecnológica a la que nos hemos referido en el apartado 2.1. En diversas pestañas disponemos de información de distinta índole, mayoritariamente en formato gráfico. En este caso, visitaremos la pestaña “Grupos de población” donde accederemos a diferentes pirámides de población. Si no disponemos de dispositivos para que el alumnado explore por sí mismo, las gráficas serán presentadas por el profesorado, en concreto mostraremos unas pirámides de población que corresponden a la distribución, según sexo y edad, del número de casos positivos, hospitalizados y no hospitalizados en España (Figuras 4, 5 y 6, respectivamente). Las pirámides muestran el porcentaje sobre el total en cada caso. Por lo tanto, sumando los porcentajes de ambos sexos en todos los grupos de edad obtendríamos el 100%.

Planteamos unas preguntas exploratorias sobre niveles básicos de comprensión del gráfico: ¿qué variable se está representando? ¿Qué representa, en cada grupo de edad y sexo, la barra correspondiente? ¿Cuánto debería sumar cada porcentaje? Tengamos en cuenta que, si podemos navegar por la página, el gráfico es interactivo de manera que al colocar el puntero del ratón sobre cada barra nos indica su valor porcentual, lo que facilita la lectura activa.



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

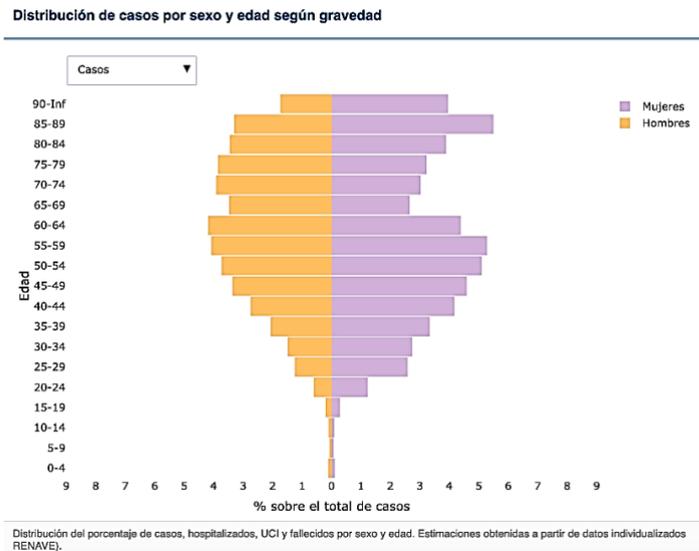


Figura 4. Pirámide de población del número de casos positivos según grupos de sexo y edad en España a 17 de mayo de 2020. Fuente: Instituto de Salud Carlos III.

A continuación, comenzamos a plantear preguntas dirigidas a identificar medidas de centralización: ¿Qué tramo de edad ha tenido más casos positivos entre los hombres? ¿Y entre las mujeres? Si tenemos acceso a la página web, y utilizando una hoja de cálculo (evítese hacerlo a mano), podemos plantear una búsqueda del intervalo mediano para el total de la población, sin discriminar entre hombres y mujeres. Nótese que, al trabajar con datos agrupados sin tener acceso a los originales, es más práctico identificar en qué intervalo se sobrepasa el 50 % de casos acumulados, empezando a contar desde el menor grupo de edad, es decir, el intervalo mediano, en lugar de localizar un valor puntual. La búsqueda del intervalo mediano la podemos hacer de varias formas. Si podemos navegar, podríamos ir identificando los porcentajes de cada barra, y sumando las barras de ambos sexos, hasta encontrar el intervalo en el que se sobrepasa por primera vez el 50 %.

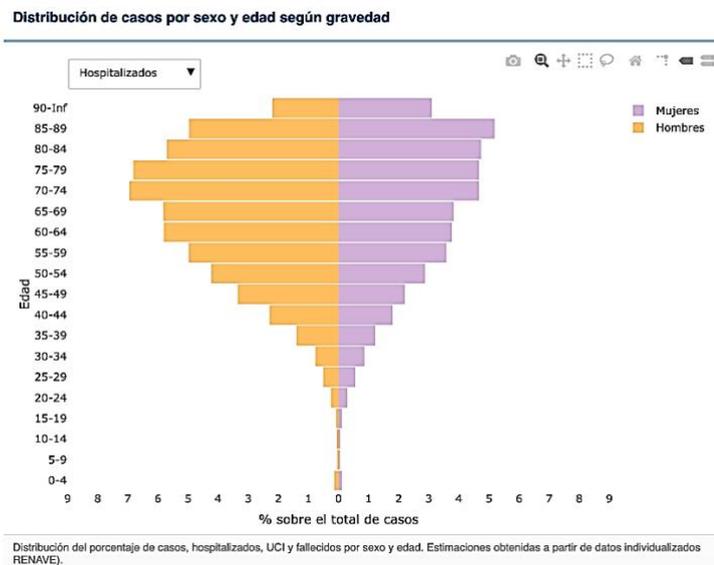


Figura 5. Pirámide de población del número de hospitalizados según grupos de sexo y edad en España a 17 de mayo de 2020. Fuente: Instituto de Salud Carlos III.

Otra forma posible de hacerlo mucho más intuitiva pasa por razonar sobre el significado gráfico de la mediana: si es aquel valor (o aquel intervalo) que ocupa la posición intermedia y las áreas de cada barra están representando el porcentaje de casos, la mediana correspondería a aquel intervalo que separa área total sombreada (de ambos colores) en dos mitades iguales. Esto nos permite experimentar incluso si no tenemos acceso a una hoja de cálculo, con una regla o un papel, que puede ser cuadriculado, representado sobre él la gráfica para facilitar el recuento.

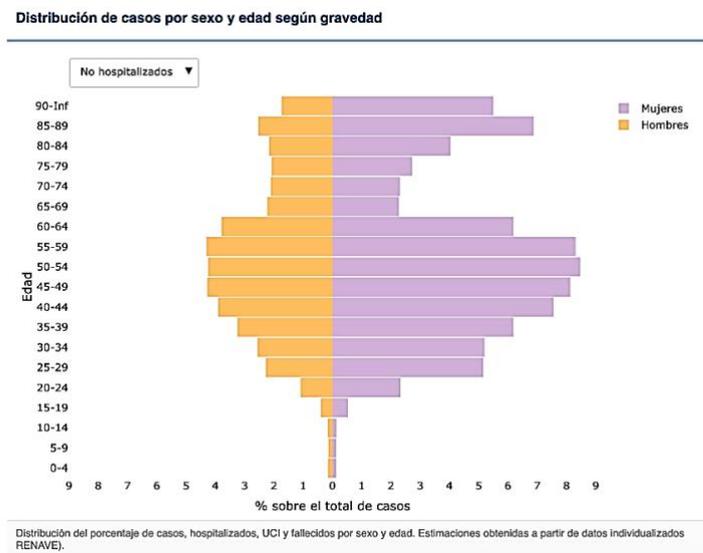


Figura 6. Pirámide de población del número de no hospitalizados según grupos de sexo y edad en España a 17 de mayo de 2020). Fuente: Instituto de Salud Carlos III.

El siguiente paso busca razonar sobre la dispersión, podemos plantear, por ejemplo, sobre la pirámide de casos positivos: ¿están más dispersos según la edad los casos en hombres o en mujeres? Como el rango es el mismo en los dos casos, surgirá la necesidad de utilizar otra medida más precisa, como la desviación típica. Pero proponemos una aproximación menos numérica y más basada en la argumentación sobre la gráfica: ¿dónde hay más diferencias de unos datos a otros? ¿Dónde hay barras de mayor tamaño más separadas entre sí? En la Figura 4 se observa con bastante claridad una mayor dispersión entre las mujeres, ya que hay más mujeres jóvenes afectadas, mientras que en los hombres los casos se concentran en edades más avanzadas. Análogamente ocurre en la Figura 6, mientras que la Figura 5 nos permitirá discutir que, si bien hay más casos de hospitalizados entre los hombres, la forma de la distribución no difiere mucho de los hombres a las mujeres.

El tercer paso que proponemos es la comparación, basada en el razonamiento conjunto de media y desviación típica: teniendo en cuenta que en la Figura 4 los datos de los hombres están más dispersos que los de las mujeres, ¿dónde será más representativa la edad media de los casos positivos? O, dicho de otro modo, ¿dónde la media representará mejor el centro de los datos? Esta pregunta nos permite utilizar, sin hacer cálculos, la idea de representatividad de la media, como valor típico, explicando la pérdida de representatividad de la media si la dispersión es mayor: los datos se parecen menos a la media.

Un paso adicional es la comparación entre distribuciones y la formulación de conjeturas. Observando las Figuras 4, 5 y 6, ¿qué crees que ocurre con las mujeres y los hombres? ¿Enferman más



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

los hombres o las mujeres? ¿Qué ocurre con la hospitalización? ¿Se comporta de igual manera que el número de casos? Es fácil observar que, a pesar de que hay más casos positivos de mujeres que de hombres (Figura 4), requieren más hospitalización los hombres que las mujeres (Figuras 5) y, además, esta diferencia se agudiza en el caso de mujeres jóvenes, que requieren menos hospitalización que los hombres (Figura 6).

Si no disponemos de dispositivos para el alumnado, podemos mostrar la Tabla 2 y preguntar ¿a cuál de las tres pirámides crees que corresponden estos datos y por qué? Se trata de un cambio de registro de representación, recorriendo en sentido contrario el habitual paso de una tabla de valores a una gráfica. Si tenemos dispositivos también lo podemos hacer, pero la identificación es más sencilla, ya que podemos determinar porcentajes barra a barra.

Tramo de edad	% sobre total H	% sobre total M	% conjunto H+M
0-4	0.14	0.11	0.25
5-9	0.03	0.03	0.06
10-14	0.06	0.05	0.11
15-19	0.08	0.11	0.19
20-24	0.25	0.29	0.54
25-29	0.51	0.56	1.07
30-34	0.77	0.86	1.63
35-40	1.4	1.21	2.61
40-44	2.3	1.8	4.1
45-49	3.35	2.2	5.55
50-54	4.23	2.88	7.11
55-59	4.98	3.58	8.56
60-64	5.81	3.77	9.58
65-69	5.82	3.83	9.65
70-74	6.96	4.67	11.63
75-79	6.82	4.66	11.48
80-84	5.7	4.74	10.44
85-89	4.95	5.19	10.14
90-Inf	2.2	3.1	0.25
Total	56.36	43.64	

Tabla 2. Distribución del porcentaje de hospitalizados (sobre el total de hospitalizados) según grupos de sexo (H=Hombre, M=Mujer) y edad en España a 17 de mayo de 2020. Fuente: Elaboración propia a partir de datos del Instituto de Salud Carlos III.

Finalmente, a la vista de la Tabla 2, y manejando exclusivamente calculadora u hoja de cálculo, plantearemos el cálculo de la edad media de los hospitalizados. Para ello, debemos discutir dos puntos importantes, a menudo pasados por alto. Como no disponemos de los datos originales, sino que están ya agrupados, debemos utilizar la marca de clase o punto medio como elemento representativo de cada intervalo de edad. Pero aquí la edad está considerada en tramos discretos, porque la respuesta habitual es un valor discreto (normalmente se responde “tengo 14 años” y no “14.7”), aunque la edad esté tratada

en el gráfico como una variable continua. Así, el punto medio del intervalo 50-54 será 52.5 y no 52. Además, que el intervalo final se denote 90-inf (entendemos que inf significa “infinito” a pesar de lo desafortunado de la expresión, ya que hubiera sido preferible 90 y más o el uso de $[90, +\infty)$, para manejar correctamente la notación intervalar), nos lleva a fijar un convenio sobre cuál sería el punto medio en ese caso, pudiendo incluso realizar pruebas con varios valores (95, 100 o 105) y ver qué efectos tendría en la media. Como su frecuencia es muy baja, no tendrá mucho efecto sobre el cálculo final, lo cual también nos da pie para discutir la influencia que sobre la media tienen los valores extremos o *outliers*.

Experiencia 4: El estudio ENE-COVID19

Nivel: 15-16 años.

Contenidos implícitos: Fases y etapas de un estudio estadístico. Representatividad de la muestra. Diagrama de dispersión e introducción a la correlación.

Descripción de la actividad: Algunos países, como España, pasada la fase más crítica de la enfermedad han comenzado a desarrollar un estudio seroepidemiológico denominado ENE-COVID19 sobre su población, con el fin de detectar qué porcentaje de personas presenta anticuerpos del virus SARS-Cov2 (el causante de la enfermedad COVID-19) y, en consecuencia, está inmunizado frente a posibles rebrotes (Ministerio de Sanidad, 2020). Utilizaremos los datos de este estudio en España para desarrollar una actividad que permita una reflexión sobre las fases y etapas de un estudio estadístico.

El Ministerio de Sanidad de España, a través del Instituto de Salud Carlos III diseñó el experimento. Dada la exposición que el alumnado ha tenido a noticias sobre las dificultades de las administraciones para realizar pruebas serológicas, podemos comenzar la discusión planteando la imposibilidad de realizar un estudio censal y la necesidad de seleccionar una muestra para analizar la proporción de población que tiene anticuerpos. A partir de esta necesidad, podemos preguntar si alguno de los alumnos o alumnas ha participado en el estudio, para comentar con los demás ¿cómo fue la llamada?, ¿qué tuvieron que hacer? y, de ese modo, tener información sobre la fase experimental del muestreo. Obviamente, si ninguno de los alumnos o alumnas ha participado en el estudio, el profesor o profesora deberá explicar cómo ha sido el proceso (la información está recogida en la web).

¿Cómo se selecciona una muestra para que sea representativa? Hay experiencias, más apropiadas a etapas anteriores, para comprobar cómo la selección aleatoria produce, en media, resultados mejores que la selección intencional (por ejemplo, midiendo la estatura de la clase con diferentes muestras intencionales seleccionadas por el alumnado o con diversas muestras aleatorias). Asumida la preferencia por la muestra aleatoria podemos reflexionar sobre cómo elegirla. Después de una discusión abierta con el alumnado, para dar valor a sus ideas al respecto, plantearemos cuál fue la selección del estudio ministerial: un muestreo bietápico estratificado.

¿De qué tipo de información sobre el total de la población dispone el Gobierno? Si hemos trabajado anteriormente con bases de datos abiertas, el alumnado conocerá el Instituto Nacional de Estadística. Quizá esté menos familiarizado con ¿qué es el censo? y ¿cómo se organiza?, pero teniendo en cuenta que el censo organiza, por ejemplo, las elecciones políticas, podemos utilizar esa información: ¿sabes dónde van a votar tus padres cuando hay elecciones? ¿Has ido allí? ¿Te has fijado en que se organiza en mesas, con unos listados? ¿Y que, según la calle o el pueblo en el que vivas, te corresponde votar en una mesa u otra? La selección del estudio se basó en el censo, seleccionando 1500 unidades censales y, dentro de cada una de ellas, 24 domicilios, lo que, multiplicado por el número de personas



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

por domicilio supuso unas 100 000 personas. Esta selección, primero de 1500 unidades censales y, luego, de 24 domicilios por cada una explica el adjetivo bietápico, ya que se ha desarrollado en dos fases, en lugar de elegir domicilios al azar sobre el total. Además, el estratificado supone que las unidades censales y los domicilios no se seleccionaron de modo que cualquiera tuviera la misma probabilidad de ser elegido, sino que se realizaron estratos (grupos homogéneos) de los hogares españoles, para que la muestra total se pareciera lo más posible a la población española.

Por otro lado, en la información ministerial, disponemos de un gráfico (Figura 7) que resulta idóneo para abordar con el alumnado la problemática de cualquier muestreo: aunque la muestra se haya seleccionado al azar, no siempre se logra contactar con todos los seleccionados, ni se les puede obligar a participar e, incluso, es posible que entre quienes estén dispuestos al final, por motivos varios, algunos no puedan participar.

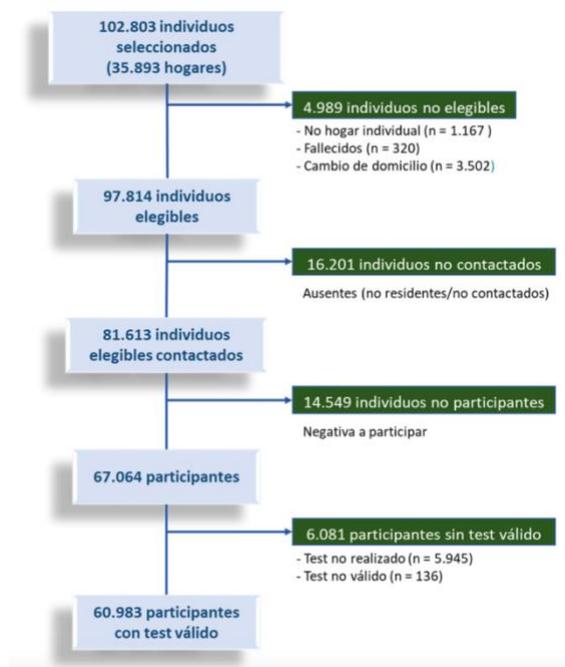


Figura 7. Proceso de reclutamiento de la muestra en el estudio ENE-COVID19. Fuente: Ministerio de Sanidad (2020).

Como vemos en la Figura 7, al final fueron casi 61 000 los participantes. ¿Qué fracción muestral supone, es decir, qué porcentaje sobre el total de la población española se ha estudiado? Esta pregunta nos permite discutir sobre la posibilidad de extraer conclusiones fiables sobre la población de 47 millones de habitantes con una muestra que no llega al 0.13 % de la población, porque lo más relevante es realizar una buena selección aleatoria. Además, podemos contrastar con el tamaño de muestra de otras encuestas importantes en España (como las del Centro de Investigaciones Sociológicas). La Figura 8 ilustra la distribución de población por rangos de edad y sexo en España y, sombreada, la distribución de la muestra seleccionada. Podemos utilizarla para valorar la representatividad de la muestra que, como se aprecia, es muy elevada.

Hombres y mujeres. Comparativa de muestra y padrón.

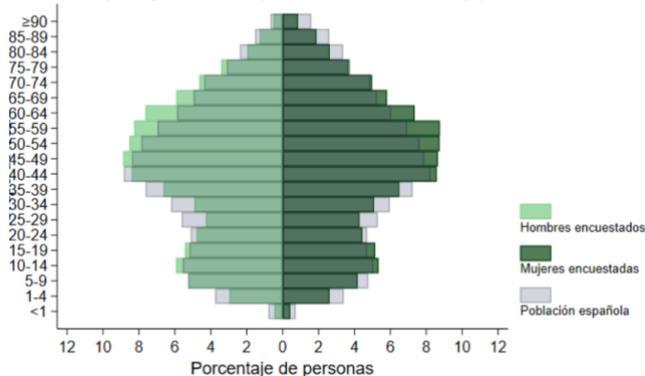


Figura 8. Comparación de la distribución por edad y sexo de la muestra del estudio ENE-COVID19 y del censo de 2019 en España. Fuente: Ministerio de Sanidad (2020).

La actividad concluye valorando la idoneidad de utilizar los valores obtenidos en la muestra (Figura 9) para estimar el verdadero porcentaje de población con anticuerpos del SARS-Cov2, reflexionando sobre las posibilidades de la estadística inferencial. También podemos utilizar la Figura 9 para plantear cuestiones respecto a la coropleta como las referidas en la Experiencia 2.

Mapa provincial de Anticuerpos IgG anti SARS-Cov2

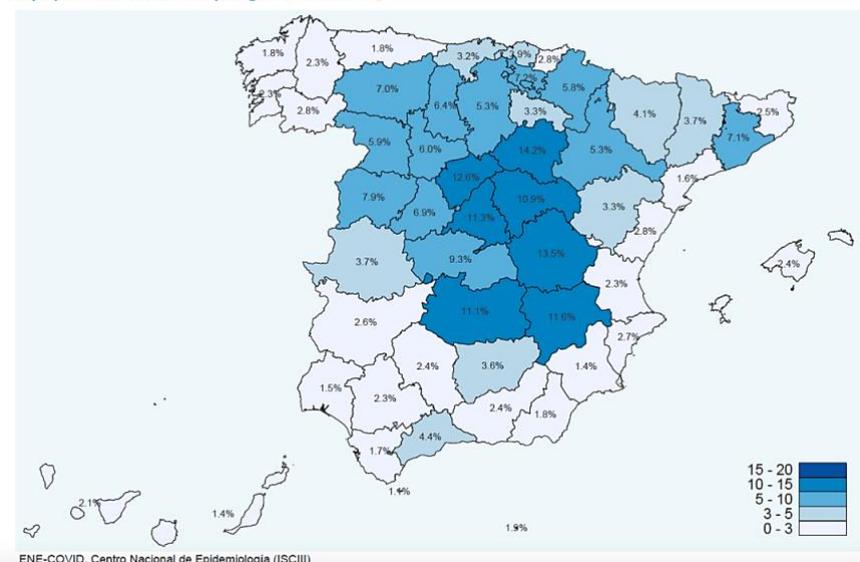


Figura 9. Coropleta de resultados por provincias del estudio ENE-COVID19. Fuente: Ministerio de Sanidad (2020).

Para completar la experiencia cambiaremos el enfoque, con el fin de abordar la correlación, para lo cual partimos de un diagrama de dispersión o nube de puntos (Figura 10). Como se observa, hay una clara correlación positiva entre ambas variables, ya que, globalmente, cuando una aumenta también lo hace la otra. A pesar de observar algunos casos anómalos que no siguen la tendencia mayoritaria, por ejemplo, el caso de los Emiratos Árabes Unidos (UAE, en la gráfica), con muchos test realizados pero una incidencia baja. De nuevo, podemos buscar una explicación con el alumnado: todos habrán



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vázquez y Á. Alsina

escuchado noticias acerca de la dificultad operativa y económica para realizar test, además de los conocimientos que han adquirido en Geografía, lo que nos permite explicar la situación de Emiratos, debido a su enorme capacidad económica (entre otros factores).

Pero la principal potencialidad didáctica de la Figura 10 es que nos permite razonar sobre las diferencias entre causalidad y correlación. La clara correlación positiva indica que cuando se realizan más test se detectan más casos positivos, pero ¿quiere esto indicar que la causa de los positivos es haber realizado el test? Dicho de otro modo ¿se contrae el virus por hacer el test? Es evidente que no, sino que lo que nos dice la gráfica es que se detectan, en general, más casos, cuando se hacen más test, frente a la política de testear únicamente a aquellas personas con síntomas (este hecho se produce por la existencia de personas infectadas pero asintomáticas).

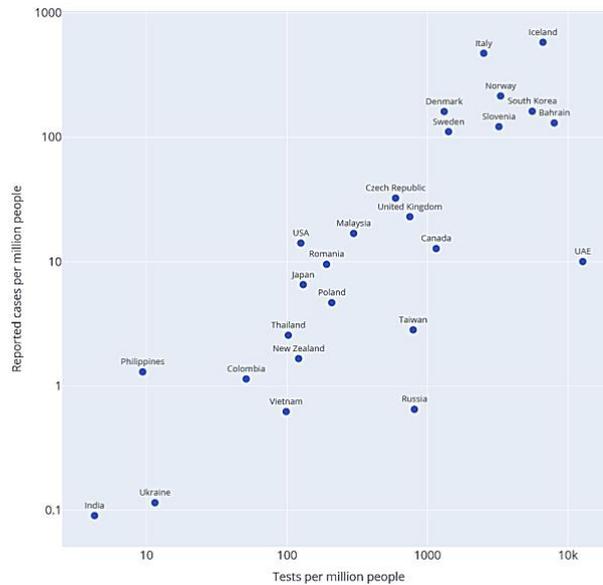
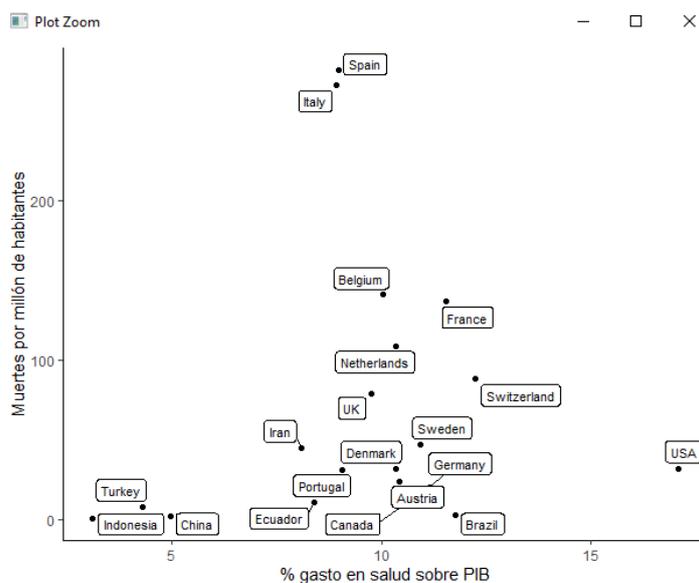


Figura 10. Diagrama de dispersión del número de test y de casos positivos por millón de habitantes en distintos países a 13 de marzo de 2020. Fuente: www.the-scientist.com.

También es interesante que mostremos al alumnado otro tipo de diagramas de dispersión sin correlaciones claras, ya que si siempre utilizamos nubes de puntos correladas podemos inducir un sesgo relativo a la correlación de modo que tiendan a apreciar correlaciones cuando no las hay. La Figura 11 nos muestra que no se aprecia una asociación clara entre el porcentaje de su PIB (Producto Interior Bruto) que cada país dedica a gasto sanitario y el número de muertes por millón de habitantes.



Fuente 11. Diagrama de dispersión del porcentaje de gasto en salud sobre PIB y número de muertes por millón de habitantes a 7 de abril de 2020. Fuente: <https://analisisydecision.es>.

Consideraciones finales

La pandemia de la COVID-19 ha comportado una crisis global, tanto social como económica, que ha venido acompañada de una gran avalancha de datos. Nunca antes las matemáticas habían estado tan presentes en el día a día de la ciudadanía: en los medios de comunicación, en los comunicados oficiales del gobierno, en las conversaciones cotidianas, etc. Ante esta situación, en este artículo se ha procurado responder a diversos interrogantes como por ejemplo ¿cómo podemos, desde la educación matemática, ayudar al alumnado a comprender la raíz, evolución y futuro de la crisis?; ¿a entender los datos científicos que se les presentan a diario?; o bien ¿a tomar decisiones informadas para proteger la propia salud y la salud comunitaria?

Para que la ciudadanía en general, y el alumnado en particular, pueda entender toda esta información (datos sociodemográficos, económicos, sanitarios y de movilidad) y tomar decisiones en consecuencia, es necesaria una alfabetización realmente interdisciplinar en la que la alfabetización estadística, probabilística y de datos tienen un papel muy relevante, pero que debe considerar también otros aspectos vinculados a los Objetivos de Desarrollo Sostenible (ODS) adoptados por la UNESCO (2017). En concreto, nos referimos por ejemplo a la salud y el bienestar (ODS3) o bien a una educación de calidad (ODS4) que tenga en cuenta, por ejemplo, la sensibilidad al trabajar estos temas con el alumnado. Aunque estemos en la etapa de Secundaria y el alumnado tenga mayor grado de madurez con respecto a Primaria, es importante tratar estas cuestiones con delicadeza, sobre todo teniendo en cuenta que puede ser un tema emocionalmente duro si han tenido una experiencia personal cercana. Desde este prisma, pues, se asume que puede ofrecerse una respuesta óptima a los interrogantes planteados desde tres dimensiones interrelacionadas: en primer lugar, desde la dimensión de la alfabetización estadística y de datos (junto con la alfabetización probabilística, que se trata en Vásquez et al, 2020); en segundo lugar, desde una dimensión metodológica que considere el trabajo en contexto y los datos reales para promover esta alfabetización; y, en tercer lugar, desde la dimensión de la educación para la sostenibilidad, que constituye un marco interpretativo y de acción.



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

Desde esta triple dimensión, en este artículo se han expuesto los conocimientos que el profesorado debería movilizar para promover la alfabetización estadística y de datos del alumnado, junto con diversas experiencias desarrolladas a partir de datos de la COVID-19 para poder trabajar estos conocimientos en contexto. En concreto, se ha partido de las recomendaciones del Proyecto GAISE (GAISE College Report ASA Revision Committee, 2016; Franklin et al. 2007), que se centran principalmente en enseñar el pensamiento estadístico como un proceso investigativo de resolución de problemas y toma de decisiones; centrarse en la comprensión conceptual; integrar datos reales con un contexto y propósito; fomentar el aprendizaje activo; o bien usar tecnología para explorar conceptos y analizar datos. Se trata, en definitiva, de promover una alfabetización estadística a partir de contextos cotidianos y de interés social junto con una alfabetización de datos a partir de herramientas como la manipulación tecnológica para manejar bases de gran tamaño, con el propósito de que progresivamente el alumnado pueda desarrollar una mirada crítica contra datos manipulados, noticias falsas, etc. En otras palabras, se trata de desarrollar el pensamiento, el razonamiento y el sentido estadístico del alumnado para que sean ciudadanos y ciudadanas bien informados y capaces de interpretar críticamente los datos. La integración de estos conocimientos junto con los derivados de la alfabetización probabilística va a permitir, además, tomar decisiones informadas tanto en el contexto de la crisis de la COVID-19 como en otras posibles crisis futuras insospechadas.

Agradecimientos

FEDER/Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación/ Proyecto EDU2017-84979-R.

FONDECYT N° 1200356 financiado por la Agencia Nacional de Investigación y Desarrollo del Gobierno de Chile.

Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación/ Proyecto TIN2017-87600-P.

Bibliografía

- Alsina, Á., Vásquez, C., Muñiz-Rodríguez, L., y Rodríguez-Muñiz, L.J. (2020). ¿Cómo promover la alfabetización estadística y probabilística en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Primaria. *Épsilon - Revista de Educación Matemática*, 104, 99-128.
- Batanero, C., Díaz, C., Contreras, J. M., y Roa, R. (2013). El sentido estadístico y su desarrollo. *NÚMEROS, Revista de Didáctica de las Matemáticas*, 83, 7-18.
- Batanero, C., y Díaz, C. (2011). *Estadística con proyectos*. Granada: Departamento de Didáctica de la Matemática.
- Burrill, G., y Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education - A joint ICMI/IASE study* (pp. 57- 69). Dordrecht: Springer.
- Castells, M. (2005). *La era de la información. Volumen 1: La sociedad red*. México, DF: Siglo XXI Editores.
- Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., y Scheaffer, R. (2005). *A Curriculum Framework for K-12 Statistics Education. GAISE Report*. American Statistical Association. Recuperado de http://www.amstat.org/education/gaise/GAISEPreK-12_Full.pdf

- GAISE College Report ASA Revision Committee (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report 2016*. Recuperado de <http://www.amstat.org/education/gaise>
- Gleason, N. W. (2018). *Higher Education in the Era of the Fourth Industrial Revolution*. Singapore: Palmgrave Macmillan.
- Hahn, C. (2015). La recherche internationale en éducation statistique: état des lieux et questions vives. *Statistique et Enseignement*, 6(2), 25-39.
- Hahn, C. (2015). La recherche internationale en éducation statistique: état des lieux et questions vives. *Statistique et Enseignement*, 6(2), 25-39.
- Marchiori, M. (2016). Little Big Data: shaping minds for the cloud. En 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (pp. 745-752). IEEE.
- Ministerio de Sanidad. España (2020). *Estudio ENE-COVID19: Primera ronda. Estudio nacional de sero-epidemiología de la infección por SARS-Cov2 en España. Informe preliminar 13 de mayo de 2020*. Recuperado de <https://www.msbs.gob.es/gabinetePrensa/notaPrensa/pdf/13.05130520204528614.pdf>
- Moore, D. y Cobb, G. (1997). Mathematics, Statistics, and Teaching. *American Mathematical Monthly*, 104, 801-823.
- Muñiz-Rodríguez, L., Rodríguez-Muñiz, L.J. y Alsina, Á. (2020). *The absence of statistical and probabilistic literacy in citizens: effects in a world in crisis*. Artículo entregado para la publicación.
- NCTM (2013). *Developing Essential Understanding of Statistics Grades 6-8*. Reston, VA: NCTM.
- OCDE (2019). *OECD Future of Education and Skills 2030: OECD Learning Compass 2030*. Paris: OCDE. Recuperado de http://www.oecd.org/education/2030-project/teaching-and-learning/learning/core-foundations/Core_Foundations_for_2030_concept_note.pdf
- Rodríguez-Muñiz, L.J., Díaz, P., y Muñiz-Rodríguez, L. (2019). Statistics and probability in the Spanish baccalaureate: Intended curriculum and implementation in textbooks. En Y. Shimizu y R. Vithal (Eds.), *24th ICMI Study Conference. School Mathematics Curriculum Reforms: Challenges, Changes and Opportunities* (pp. 413-420). Tsukuba, Japan: ICMI, University of Tsukuba.
- Schild, M. (2004). Information literacy, statistical literacy and data literacy. *Iassist Quarterly (IQ)*, 28(2/3), 6-11.
- UNESCO. (2017). *Educación para los objetivos de desarrollo sostenible: objetivos de aprendizaje*. Francia: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

Luis J. Rodríguez-Muñiz. Departamento de Estadística e I.O. y Didáctica de la Matemática. Facultad de Geología, C/ Jesús Arias de Velasco s/n. 33005 Oviedo (Asturias).
Profesor Titular de Universidad en Didáctica de la Matemática, es autor de más de 50 artículos y capítulos de libro. Su investigación reciente se focaliza en la educación estadística y la formación de profesorado de Primaria y Secundaria. Ha participado en numerosos proyectos de investigación y también en contratos con administraciones públicas sobre evaluación del sistema educativo. Preside, desde 2018, la Comisión de Educación de la Real Sociedad Matemática Española (RSME), de cuya Junta Directiva es miembro desde 2020. También es socio de la Sociedad Española de Investigación en Educación Matemática (SEIEM). Además, es miembro, como experto de reconocido prestigio, del Consejo de Asturias de la Formación Profesional, y de la Comisión Asesora de Evaluación de Enseñanzas e Instituciones de ANECA. Email: luisj@uniovi.es



¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria

L.J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, C. Vásquez y Á. Alsina

Laura Muñiz-Rodríguez. Departamento de Estadística e I.O. y Didáctica de la Matemática, Facultad de Geología, C/ Jesús Arias de Velasco, s/n, 33005 Oviedo (Asturias).

Licenciada en Matemáticas y Doctora en el Programa de Matemáticas y Estadística por la Universidad de Oviedo y Doctora en Ciencias de la Educación por la Universidad de Gante (Bélgica). Profesora Ayudante Doctora en el Departamento de Estadística e I.O. y Didáctica de la Matemática de la Universidad de Oviedo. Sus publicaciones se han centrado en el campo de la formación inicial del futuro profesorado de matemáticas en Educación Primaria y Secundaria, el uso de juegos como recurso didáctico para la enseñanza y el aprendizaje de las matemáticas, y el proceso de retroalimentación en el aula. Email: munizlaura@uniovi.es

Claudia Vásquez. Departamento de Didáctica de la Matemática. Campus Villarrica, O'Higgins 501, Villarrica, Chile.

Licenciada en Matemática por la Pontificia Universidad Católica de Chile, y Doctor en Ciencias de la Educación por la Universidad de Girona (España).

Actualmente es Profesor Asociado de la Pontificia Universidad Católica de Chile. Sus líneas de investigación se centran en la enseñanza y aprendizaje de la estadística y la probabilidad, y en la formación del profesorado. Ha participado en proyectos de investigación sobre didáctica de la matemática, didáctica de la probabilidad y la estadística, y formación del profesorado. Email: cavasque@uc.cl

Ángel Alsina. Departamento de Didácticas Específicas, Área de Didáctica de las Matemáticas. Facultad de Educación y Psicología, Plaça Sant Domènec, 9, 17004 Girona (Catalunya).

Catedrático de Didáctica de las Matemáticas en la Universidad de Girona. Sus líneas de investigación están centradas en la enseñanza y el aprendizaje de las matemáticas en las primeras edades y en la formación del profesorado de matemáticas. Ha publicado numerosos artículos científicos y libros sobre cuestiones de educación matemática, y ha llevado a cabo múltiples actividades de formación permanente del profesorado de matemáticas en España y en América Latina. Email: angel.alsina@udg.edu

Notas:

¹ https://es.wikipedia.org/wiki/Bot_social