RECHIEM

REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA

V.13
NÚMERO ESPECIAL
- 4 -

LA EDUCACIÓN ESTADÍSTICA EN EL AULA ESCOLAR CHILENA

sochiem

ÍNDICE

sochiem

Sociedad Chilena de Educación Matemática
Revista Chilena de Educación Matemática
ISSN 2452-5448 Versión en línea
Chile

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

EDITORIAL

https://doi.org/10.46219/rechiem.v13i4.100

Editora invitada
Dra. Claudia Vásquez Ortiz
cavasque@uc.cl
Pontificia Universidad Católica de Chile

La Educación Estadística en el aula escolar chilena

Prefacio

En la actualidad cobra gran relevancia lo señalado por H. G. Wells a comienzos del siglo XX referido a que "el comprender promedios, máximos y mínimos algún día será tan necesario para una ciudadanía eficiente como lo es hoy la habilidad de leer y escribir" (Wells, 1911, p. 204)[1], pues estamos insertos en un mundo en que los datos abundan, y es necesario saber interpretarlos para una toma de decisiones informada. Por tanto, como profesores y formadores de profesores, hoy más que nunca estamos desafiados a formar ciudadanos con un pensamiento crítico, capaces de leer y analizar datos de manera que puedan reconocer cuando la información es comunicada de manera engañosa o inapropiada (OCDE, 2019)[2].

Chile no está ajeno a este desafío, es más, la inclusión de temas de estadística y probabilidad en el currículo escolar goza de cierta trayectoria y está presente al menos desde mediados de la década de los 60's, por ejemplo, en el currículo de Educación Básica (Figura 1). Tal es el caso de los libros de texto y programas de estudio de dicha época en que en los últimos cursos de Educación Básica se observa el estudio de promedios de datos, gráficos de barras, pictogramas, azar y probabilidades, entre otros.

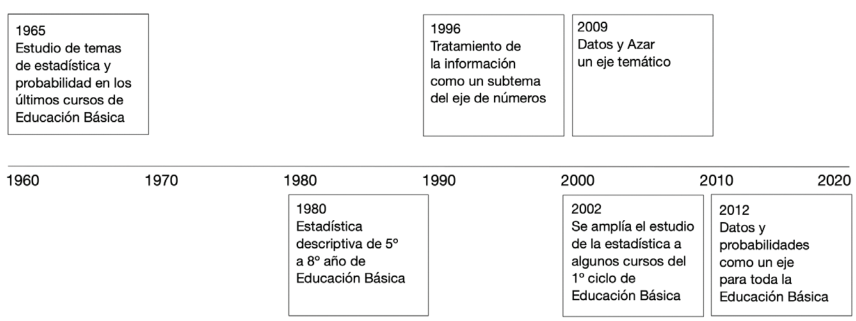


Figura 1. Una cronología de los cambios en el currículo escolar de matemática respecto de la enseñanza de la estadística y la probabilidad. Fuente: Elaboración propia.

De este modo, en 1980 se promulga el Decreto 4002 con fuerza de Ley que fija nuevos objetivos, planes y programas para la Educación Básica, en el cual, respecto a la asignatura de matemática se especifica que para el 2º ciclo de Educación Básica (de 5º a 8º grado) los estudiantes deben conocer y aplicar elementos de estadística descriptiva. Posteriormente, en el año 1996, a partir del Decreto 40 de la Ley N° 18.962 Orgánica Constitucional de Enseñanza, se modifican los objetivos y contenidos para la Educación Básica. En el caso de la asignatura de matemática, se observa una agrupación de los contenidos en cuatro ejes temáticos: números, operaciones aritméticas, formas y espacio, y resolución de problemas. Así, en el eje de números se encuentra el subtema llamado "tratamiento de la información" el cual está presente en 3º, 4º, 7º y 8º grado de Educación Básica. A través de este subtema se pretende que los estudiantes recolecten, interpreten y analicen datos provenientes de situaciones del entorno local, regional y nacional y comuniquen sus resultados por medio de la comparación de datos, promedios y valor más frecuente y medidas de tendencia central, además de utilizar distintos tipos de representaciones estadísticas como tablas de frecuencias relativas, gráficos circulares entre otras.

[1] *Wells, H. G. (1911). Mankind in the Making. 5th ed. Chapman and Hall. [originally published in 1903]*
[2] *Organización para la Cooperación y el Desarrollo Económicos [OCDE]. (2019). OECD Future of Education and Skills 2030: OECD Learning Compass 2030. Autor.*

Posteriormente, en el año 2002, se modifican los objetivos y contenidos para la asignatura de matemática donde en 3º y 4º grado se incluye la interpretación y organización numérica en tablas y gráficos de barras, para después, a partir del subtema "tratamiento de la información" ligado al eje de números, los estudiantes de 6º, 7º y 8º año de Educación Básica se centren en el aprendizaje de diversas temáticas vinculadas a la estadística (medidas de tendencia central, tablas y gráficos estadísticos).

Mas recientemente, en el año 2009, se realiza un ajuste al currículo, el cual organiza la asignatura de matemática en torno a los ejes curriculares de: números, álgebra, geometría y datos y azar. Así, a través de este último eje, por primera vez, la estadística y la probabilidad es situada en el currículo al mismo nivel que los otros temas, con el propósito de introducir a los estudiantes en el tratamiento de datos y modelos para el razonamiento en situaciones de incerteza. Para ello, se plantea abordar temas de estadística desde 1º año básico (recolección, interpretación y análisis de datos) y temas de probabilidad desde 5º año básico (uso de lenguaje asociado a la probabilidad, experimentos aleatorios, cálculo de probabilidades).

Consecutivamente, en el año 2012, con el propósito de adaptar el currículo a los requerimientos del mundo actual, se definen las nuevas Bases Curriculares para la Educación Básica. En dichas bases, la asignatura de matemática se estructura en torno a cinco ejes temáticos (números y operaciones, patrones y álgebra, geometría, medición, datos y probabilidades). Por medio del eje de datos y probabilidades, se aborda el estudio de estos temas de manera gradual y continua desde 1º año de Educación Básica, con la finalidad de que "todos los estudiantes registren, clasifiquen y lean información dispuesta en tablas y gráficos, y que se inicien en temas relacionados con las probabilidades. Estos conocimientos les permitirán reconocer gráficos y tablas en su vida cotidiana. Para lograr este aprendizaje, es necesario que conozcan y apliquen encuestas y cuestionarios por medio de la formulación de preguntas relevantes, basadas en sus experiencias e intereses, y después registren lo obtenido y hagan predicciones a partir de ellos" (MINEDUC, 2012, p. 5)[3].

Como hemos observado la estadística y la probabilidad han estado presentes en currículo escolar chileno de matemática desde hace varias décadas. Todo esto con el propósito de que los estudiantes al finalizar su etapa escolar cuenten con las competencias necesarias para desenvolverse como ciudadanos críticos, tanto en el mundo laboral, cotidiano y en la universidad. No obstante, su enseñanza sigue constituyendo un reto para el profesorado, especialmente para aquellos docentes que se desempeñan en los niveles de Educación Básica, quienes, en muchos casos, no han recibido una preparación durante su formación inicial para enseñar estos temas, y, por tanto, muchas veces evitan su enseñanza o bien la sitúan como uno de los últimos temas a tratar en los cursos, con un enfoque muy centrado en lo algorítmico.

Por consiguiente, surge la necesidad de promover y reforzar la investigación en Educación Estadística para contribuir a la generación de conocimiento en el área y para mejorar la preparación del alumnado y el profesorado. Aún más, si consideramos los hechos recientemente acontecidos en los medios de comunicación nacional e internacional que dan a conocer errores, por ejemplo, en la construcción de gráficos estadísticos; estos no hacen más que reflejar la urgente necesidad de educar a los ciudadanos en estadística y probabilidad, desarrollando las competencias para comprender e interpretar el mundo desde la primera infancia, esto es, desde el nivel de Educación Parvularia. Para ello, es importante prestar atención a la investigación desarrollada en Educación Estadística, en especial, aquella vinculada al aula escolar, pues estas investigaciones son todavía escasas y poco se conoce respecto de cómo abordar la enseñanza de la estadística y la probabilidad en los distintos niveles educativos, aun más en el contexto chileno.

En consideración a lo anteriormente expuesto, con este número especial queremos contribuir en la difusión de investigaciones sobre estadística y probabilidad en el aula escolar chilena, concretamente, en los niveles escolares de Educación Parvularia, Educación Básica y Educación Media, así como, en la formación inicial y continua del profesorado de dichos niveles educativos. De este modo, esperamos brindar un panorama de los diversos tipos de investigación que se están llevando a cabo sobre estos temas en el aula escolar chilena a nivel nacional e internacional.

A continuación, se presentan las nueve contribuciones que conforman este número especial: cuatro publicaciones corresponden a experiencias de aula, tres a investigaciones vinculadas al conocimiento estadístico y otras dos, al análisis de libros de texto.

El primer artículo, "Niveles de lectura y contextos en las actividades sobre tablas estadísticas en libros de texto chilenos y españoles", escrito por los investigadores *Jocelyn D. Pallauta, María M. Gea, Carmen Batanero y Pedro Arteaga*, presenta un análisis, según los niveles de lectura, de las tareas planteadas en torno a las tablas estadísticas en una muestra de libros de texto chilenos y españoles. Asimismo, se examina el contexto de dichas tareas, de acuerdo con los lineamientos de los estudios PISA.

[3] *Ministerio de Educación de Chile [MINEDUC]. (2012). Bases Curriculares 2012: Educación Básica Matemática. Unidad de Currículum y Evaluación. Autor.*

En el segundo artículo, los investigadores *Pedro Vidal-Szabó y Soledad Estrella* presentan el estudio "Conocimiento estadístico especializado en profesores de Educación Básica, basado en la taxonomía SOLO". Este trabajo explora los conocimientos y niveles de comprensión que manifiesta un grupo de profesores de Educación Básica sobre los conceptos de variable estadística, dato, información estadística y contexto de los datos.

El tercer artículo, a cargo de autores *Hugo Alvarado Martinez, Sergio Tapia Munoz, Lidia Retamal y Liliana Tauber,* presenta la investigación "Explorando las nociones probabilísticas informales en estudiantes de educación básica". En dicho estudio se analizan las intuiciones y heurísticas sobre la probabilidad en estudiantes de 12 a 13 años de edad, con el propósito de favorecer un acercamiento a la comprensión de los significados intuitivo y clásico de la probabilidad.

En el cuarto artículo, titulado "Conocimiento sobre el muestreo en estudiantes chilenos al término de la educación escolar", los autores *Karen Ruíz-Reyes, Felipe Ruz, Elena Molina-Portillo y José M. Contreras* indagan respecto de la comprensión sobre los métodos de muestreo en estudiantes chilenos que cursan su último año de escolaridad. Para ello, analizan las producciones de los estudiantes frente a un ítem de respuesta abierta sobre la identificación del sesgo de un método de muestreo.

En el quinto artículo, las autoras *Elisabeth Ramos-Rodríguez, Natalia Alvarado-Garcés*, *Patricia Vásquez y Andrea Vergara* presentan los resultados de implementación de la experiencia de aula "Medidas de tendencia central y dispersión miradas desde un deporte típico chileno y la modelación estadística", con estudiantes de 15 a 16 años de edad, que considera la modelación estadística para la enseñanza de las medidas de tendencia central y dispersión, a partir del contexto que provee un deporte típico chileno.

En el sexto artículo, titulado "El significado de la media, mediana y moda en textos escolares de séptimo básico", los investigadores *Jaime I. García-García, Ingrid B. Urrutia Leiva, Sebastian H. Vasquez Chicao y Elizabeth Hernández Arredondo* presentan algunas sugerencias respecto de los elementos a considerar al momento de enseñar las medidas de tendencia central en el aula escolar. Para ello, se fundamentan en el significado de la media, mediana y moda en una muestra de textos escolares de séptimo grado de Educación Básica.

El séptimo artículo, a cargo de los autores *Fabiola Arévalo-Meneses y Julio Manzanares*, presenta el estudio "Preguntas elaboradas por profesores para el estudio de gráficos de barras estadísticos: los niveles de lectura que se identifican en sus propuestas". En dicha investigación se indaga en el nivel de lectura de gráficos estadísticos presentes en las preguntas elaboradas por profesores de matemática en Chile.

En el octavo artículo, titulado "Razonamiento estadístico en el contexto Covid-19: una propuesta basada en GeoGebra", los autores *Manuel González-Navarrete e Iván Maldonado-Carrasco*, presentan una propuesta didáctica centrada en el análisis exploratorio de datos asociados a la evolución de la pandemia en distintas comunas de Chile. Para ello, han diseñado un conjunto de actividades para realizar un análisis de regresión utilizando el software GeoGebra.

En el último artículo, "Interpretación y comprensión de gráficos estadísticos por profesores de matemáticas en formación", a cargo de los investigadores *Nicolás Sánchez, Elizabeth Toro Barbieri y Daniela Araya Bastias*, se explora la comprensión gráfica de estudiantes de la carrera de Pedagogía en Matemática, en una actividad asociada a la esperanza de vida. Los resultados evidencian que gran parte de los futuros profesores se encuentran en un nivel inicial de compresión gráfica.

Finalmente, agradecer al Equipo Editorial de la Revista Chilena de Educación Matemática por el apoyo y espacio brindado para la elaboración y difusión de este número especial, a los autores por sus contribuciones, así como a los referees por su participación en el proceso de revisión de estas.

Se invita a los lectores a conocer más acerca de las perspectivas actuales de la investigación en Educación Estadística en el aula escolar chilena, y a ser partícipes reflexivos y activos de las mejoras en esta materia, en todos los niveles educativos, en la formación inicial de profesores de matemática y a lo largo del desarrollo permanente de los mismos.

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

https://doi.org/10.46219/rechiem.v13i4.74

NIVELES DE LECTURA Y CONTEXTOS EN LAS ACTIVIDADES SOBRE TABLAS ESTADÍSTICAS EN LIBROS DE TEXTO CHILENOS Y ESPAÑOLES

READING LEVELS AND CONTEXT IN ACTIVITIES ON STATISTICAL TABLES IN CHILEAN AND SPANISH TEXTBOOKS

Jocelyn D. Pallauta
jocelyndiaz@correo.ugr.es
Universidad de Granada, Granada, España

Carmen Batanero
batanero@ugr.es
Universidad de Granada, Granada, España

María M. Gea
mmgea@ugr.es
Universidad de Granada, Granada, España

Pedro Arteaga
parteaga@ugr.es
Universidad de Granada, Granada, España

RESUMEN

Las tablas estadísticas son frecuentemente utilizadas para presentar y analizar la información, lo que lleva a incluir su enseñanza en la educación obligatoria. En este trabajo se analizan las actividades planteadas en torno a las tablas estadísticas en una muestra de libros de texto chilenos y españoles, dirigidos a los niveles 7º y 8º básico chilenos (1º y 2º de secundaria españoles). Las actividades se clasifican según sus niveles de lectura. También se analiza el contexto de las tareas, siguiendo los propuestos en los estudios PISA, en los que, adicionalmente, se incorporan las categorías de experimentos aleatorios y actividades sin contexto. Los resultados muestran un ligero incremento hacia los niveles sofisticados de lectura, conforme se progresa de curso, tanto en los libros chilenos como españoles, mientras que el contexto que aparece con gran fuerza en la muestra es el de tipo personal.

PALABRAS CLAVE:
Libros de texto; tablas estadísticas; niveles de lectura; contexto.

ABSTRACT

Statistical tables are frequently used to present and analyze information and are therefore included in compulsory education. In this paper, we analyze the activities proposed on statistical tables in a sample of Chilean and Spanish textbooks directed to 7th and 8th grades in Chile (1st and 2nd levels of Spanish secondary school). The activities are classified according to their reading levels. The context of the tasks is also analyzed, following those proposed in the PISA studies, to which random experiments and activities without context are added. The results show a slight increase towards the s'mores sophisticated levels of reading, as the course progresses, both in the Chilean and Spanish textbooks, while the most frequent context in the sample is the personal context.

KEYWORDS:
Textbooks; statistical tables; reading levels; context.

Recibido: 10 de Mayo de 2021, Aceptado: 10 de Agosto de 2021

1. Introducción

En los medios de comunicación, así como en el trabajo profesional, se encuentran con frecuencia tablas estadísticas que cualquier ciudadano competente necesita interpretar para comprender su información y tomar buenas decisiones, especialmente en el contexto de incertidumbre que la sociedad ha debido enfrentar por la pandemia generada por la COVID-19 (Rodríguez-Muñiz et al., 2020). Por este motivo, la competencia con las tablas estadísticas la entendemos como parte de la alfabetización estadística, o capacidad para comprender e interpretar la información estadística y razonar sobre ella (Gal, 2019; Gould, 2017).

Este trabajo se centra en los libros de texto, que constituyen un recurso didáctico valioso, al ser un puente entre las directrices curriculares y el trabajo del aula (Herbel, 2007). Esta importancia se ve reflejada en el aumento de la investigación sobre el libro de texto de matemáticas (Pepin y Gueudet, 2020), ya que se ha mostrado que este recurso tiene una influencia directa en el aprendizaje del estudiante (Reys et al., 2004).

El estudio de la tabla estadística es importante no solo por su relevancia en la presentación de la información en diferentes áreas del conocimiento, sino también debido a que su comprensión es básica para el aprendizaje de otros temas de la estadística (Estrella y Estrella, 2020). Su enseñanza se incluye en Chile desde los primeros cursos de educación básica (Ministerio de Educación de Chile [MINEDUC], 2018): de 1º a 3º básico se propone el trabajo con tablas de conteo, como un medio para organizar datos que pueden emanar del entorno cercano al estudiante. A partir de 4º básico se incluyen las tablas con frecuencias absolutas para registrar, entre otros, los resultados obtenidos de experimentos aleatorios. En 5º aparecen las tablas de datos de doble entrada, mientras que en los siguientes cursos se incorporan las tablas con diferentes tipos de frecuencias (relativas, porcentuales), que son utilizadas para representar información acerca de una muestra, o para presentar los resultados de experimentos aleatorios. En 6º y 7º el uso de las tablas estadísticas se asocia al estudio de la probabilidad frecuencial, y en 8º se hace hincapié en que los estudiantes evalúen la forma en que los datos están presentados (MINEDUC, 2015).

En las directrices curriculares de España (Ministerio de Educación, Cultura y Deporte de España [MECD], 2014), a lo largo de la educación primaria (6 a 12 años) se trabaja con tablas estadística para la recogida y clasificación de datos cualitativos y cuantitativos, así como la construcción de tablas de frecuencias absolutas y relativas. En la enseñanza secundaria (MECD, 2015), en los dos primeros cursos (12 a 14 años) se profundiza en el estudio de las tablas con frecuencias absolutas y relativas, que se usan junto a las medidas de centralización y el rango para el estudio de la probabilidad desde el enfoque frecuencial.

Basándonos en la relevancia de la tabla estadística en los lineamientos curriculares de ambos países, y la escasa literatura al respecto (Estrella y Estrella, 2020), nos planteamos como pregunta de investigación si influye el contexto educativo del país en el tratamiento del nivel de lectura y el contexto con que se plantea el estudio de la tabla estadística. Para responder a esta cuestión, en el presente trabajo se analizan las actividades planteadas en torno a las tablas estadísticas en una muestra de libros de texto chilenos y españoles en los niveles 7º y 8º básico chilenos (1º y 2º de secundaria españoles), considerando las variables niveles de lectura y el contexto en que es planteada la tabla estadística, lo que nos permitirá comparar estas variables en dos contextos educativos diferentes.

En los siguientes apartados se describen los fundamentos y antecedentes que sustentan el presente trabajo, posteriormente se detalla la metodología utilizada, así como los resultados del estudio junto a las principales conclusiones.

2. Fundamentos

El trabajo se apoya en los niveles de lectura descritos para los gráficos y tablas estadísticas por Curcio y sus colaboradores (Friel, et al., 2001), así como en el papel del contexto en el desarrollo de la actividad matemática y también en algunos antecedentes.

2.1 Niveles de lectura de gráficos y tablas estadísticas

La tabla estadística es un objeto semiótico complejo (Pallauta y Arteaga, 2021), por lo que a partir de ella pueden plantearse preguntas de diferente dificultad respecto a la información que recoge. En este sentido, en el presente trabajo se utilizan los niveles de lectura establecidos por Curcio (1989) para gráficos estadísticos y ampliados luego por Shaughnessy et al. (1996) con un cuarto nivel, que finalmente fue integrado por Friel et al. (2001). Estos niveles son aplicables y también se han utilizado para analizar tablas estadísticas (Díaz-Levicoy et al., 2018; García-García et al., 2019; Pallauta et al., 2021). Se describen resumidamente a continuación estos niveles, que se detallarán en la sección de resultados, junto con ejemplos de su uso en las actividades analizadas.

- *Nivel 1. Leer los datos,* consiste en la lectura literal de la información expuesta en un gráfico o tabla, y corresponde al nivel más básico de comprensión. Este nivel implica identificar aspectos elementales de la tabla, con el objeto de responder a preguntas en que la información esté expuesta de manera explícita en ella, por ejemplo, la lectura del título o de una celda de la tabla.

- *Nivel 2. Leer dentro de los datos,* cuando para responder a alguna cuestión, además de obtener la información anterior, es necesario realizar operaciones aritméticas, comparaciones con los datos expuestos en la tabla, o encontrar relaciones entre los datos. Por ejemplo, determinar la moda en una tabla de distribución o calcular el total que corresponde a un rango de valores en la tabla.

- *Nivel 3. Leer más allá de los datos*, involucra una lectura de información que no viene directamente representada en la tabla, mediante inferencias o razonamientos sobre los datos. Para ello se requiere extrapolar, predecir o inferir a partir de los datos presentes en la tabla. Por ejemplo, si la variable es numérica, extrapolar la frecuencia de un valor no representado, o en el caso de una tabla de doble entrada, determinar si existe asociación entre las variables representadas.

- *Nivel 4. Leer detrás de los datos*, es realizar una valoración crítica de las conclusiones, o de la recogida y tratamiento de la información expuesta en una tabla o gráfico. Por ejemplo, juzgar la fiabilidad de la muestra o la posible manipulación de los datos. En este nivel se requiere un conocimiento del contexto en que se recogieron los datos presentados.

2.2 El contexto en matemáticas y estadística

La contextualización del conocimiento matemático tiene un rol importante en la manera en que el significado de cada objeto matemático es construido, aprendido, activado y transformado (Font, 2007). El contexto tiene una fuerte influencia en la comprensión de un problema y la elección de un método para su solución, pero también este puede generar dificultades para caracterizar un problema matemático (Wijaya et al., 2014).

Actualmente existe un interés creciente por la importancia del contexto en educación matemática, puesto que la comprensión y resolución de problemas contextualizados forma parte de la cultura científica, es decir, del conjunto de conocimientos y competencias que necesita el ciudadano para desenvolverse en situaciones en que requiere del conocimiento científico (Rosales et al., 2020). Los contextos atraen la atención de los estudiantes y conectan lo que se aprende en la escuela con la sociedad externa a la misma, motivándole para seguir aprendiendo (Sanmartí y Márquez, 2017).

Diversas investigaciones han puesto en evidencia la importancia de proponer a los estudiantes tareas matemáticas que reflejen la realidad (Van den Heuvel-Panhuizen, 2005), es decir, situaciones que presenten aspectos del diario vivir, pues a pesar de ser de entorno cercano al aprendiz se pueden manifestar dificultades en su resolución (Blum et al., 2007; Wijaya et al., 2014). Es así como los estudios internacionales de evaluación PISA (Organización para la Cooperación y el Desarrollo Económicos [OECD], 2019) recogen este aspecto, pues se considera que la cultura matemática influye en la capacidad de formular y usar la matemática en una variedad de contextos. Los ítems matemáticos propuestos en dichas pruebas utilizan situaciones del mundo real, incluso cuando sean hipotéticas (Turner, 2006), que se clasifican en la forma siguiente:

- *Contexto personal.* Se trata de contextos cercanos a la experiencia del estudiante, sobre temáticas cotidianas de su entorno inmediato.

- *Contexto laboral.* Se presentan situaciones relacionadas con el mundo laboral y profesional en diversos ámbitos.

- *Contexto público.* Se trata de situaciones en el entorno local o en contextos más amplios, a las que es posible acceder a través de los diferentes medios de comunicación.

- *Contexto científico.* Incluye aplicaciones de las matemáticas más abstractas y complejas que los anteriores contextos, asociadas a procesos científicos o tecnológicos.

En este documento, se señala que el contexto moviliza diversos procesos cognitivos en el estudiante cuando resuelve un problema, por lo que en este estudio también se aborda esta variable, que ha sido valorada en algunos análisis de las actividades sobre tablas estadísticas propuestas en libros de texto (Díaz-Levicoy et al., 2015, 2018; García-García et al., 2019; Pallauta et al., 2021).

2.3 Antecedentes

La investigación cuenta con antecedentes de estudios que han analizado, entre otras variables, los niveles de lectura o los contextos de las actividades propuestas en relación a las tablas estadísticas en libros de texto dirigidos a diferentes niveles educativos.

Los principales antecedentes de análisis de tablas estadísticas en los libros de texto chilenos dirigidos a los primeros cursos de educación primaria son aportados por Díaz-Levicoy y colaboradores (Díaz-Levicoy et al., 2015, 2018). En el primero de estos trabajos, los autores analizaron las tablas estadísticas en cuatro libros de texto de 1º y 2º de educación primaria de dos editoriales chilenas. Entre las variables consideraron los niveles de lectura propuestos por Curcio y colaboradores (Curcio, 1989; Friel et al., 2001), observando la presencia de solo los dos primeros niveles de lectura, *leer los datos* (44,8%) y *leer entre los datos* (55,2%), mientras que el contexto mayormente utilizado, entre los propuestos en las pruebas PISA (OECD, 2019), fue el de tipo personal (69%), seguido del público (19%). En el análisis realizado a otros tres libros de texto dirigidos a 3º curso de primaria (Díaz-Levicoy et al., 2018), se observó que el contexto personal apareció con más frecuencia (83,5%), en este caso los niveles de lectura no fueron considerados.

García-García et al. (2019) analizan doce libros de texto mexicanos de educación primaria (1º a 6º) de dos editoriales diferentes, considerando entre las variables los niveles de lectura y el tipo de contexto en que se presenta la tabla. Aunque dicho estudio entrega resultados generales, se pudo apreciar la alta presencia de los dos primeros niveles de lectura *(leer datos y leer dentro de los datos)* en la mayoría de los textos, el nivel *leer más allá de los datos* estuvo ausente en las actividades propuestas, mientras que el nivel *leer detrás de los datos* solo estuvo presente de 2º a 6º curso en una de las editoriales analizadas. Respecto al contexto, de los considerados por PISA (OECD, 2019), apareció con mayor fuerza el de tipo personal.

Un antecedente inmediato (Pallauta et al., 2021) es el análisis realizado a las tareas escolares relacionadas con las tablas estadísticas presentadas en textos escolares chilenos de 5º a 8º básico (10 a 13 años) de dos editoriales que se distribuyen gratuitamente en el sistema público y subvencionado de Chile. El análisis de contenido de las tareas contempló, entre otras variables, el contexto, de los planteados en el informe PISA, y el nivel de lectura requerido para desarrollar cada actividad. Se observó una predominancia del nivel de lectura L2, *leer dentro de los datos* (77,4%), seguido del nivel L1, *leer los datos* (19,4%). En cuanto a los contextos, el que apareció con mayor fuerza fue el de tipo personal (41,7%), seguido del contexto ocupacional o laboral (16,8%); los autores añaden los experimentos aleatorios como nuevo contexto por su presencia significativa (16,1%) en las actividades planteadas.

En este trabajo se analiza el nivel de lectura junto al contexto en que se presenta la tabla estadística en una nueva muestra de libros de texto chilenos y otra de textos españoles, los primeros dirigidos a 7º y 8º curso de enseñanza básica, y los segundos a 1º y 2º de educación secundaria (12 a 14 años), con el objeto de comparar los resultados de ambos países. Puesto que no tenemos suficientes antecedentes del tratamiento de las tablas estadísticas en estos niveles educativos y en textos españoles, este trabajo aporta información original.

3. Metodología

La muestra se compone de seis libros de texto chilenos dirigidos a 7º y 8º básico (10 y 11 años) y seis españoles de 1º y 2º de educación secundaria, tres por cada nivel escolar analizado. La muestra de textos seleccionada (detallada en el Apéndice) es intencional, y se ajusta a criterios enmarcados por el objetivo del presente estudio (Corral et al., 2015). La codificación empleada para referirse a cada texto al incluir los ejemplos mostrados en el trabajo denota el tipo de enseñanza de cada país. Así, por ejemplo, el código EGB (Educación General Básica) se refiere a los libros de texto chilenos y cuando se indica ESO (Educación Secundaria Obligatoria) corresponde a cada uno de los textos españoles.

La muestra de libros de texto chilenos forma parte de tres series diferentes: una subvencionada por el Ministerio de Educación en Chile, publicada por dos editoriales diferentes en edición especial, SM (7º básico) y Santillana (8º básico), por otro lado, hay textos editados por SM y Santillana directamente. Estos textos son distribuidos de forma gratuita a los estudiantes del sistema público y concertado. Además, en la muestra se incluyen otros textos editados por Santillana y SM que siguen las recomendaciones oficiales (MINEDUC, 2015), los cuales se encuentran vigentes y disponibles en el mercado. El criterio de la selección de estos textos fue su uso generalizado en las escuelas concertadas y privadas, además de haber sido publicados por editoriales de tradición en el país. La muestra de libros de texto españoles ha sido publicada por las editoriales Anaya, Edelvives y Santillana, que son ampliamente utilizadas en las diferentes comunidades autónomas de España, y atienden al marco curricular vigente (MECD, 2015).

En los libros de texto se analizaron las actividades o ejemplos que requerían el uso o construcción de una tabla estadística. En la Tabla 1 se presenta la distribución de las situaciones analizadas por curso y país. En total, se revisó un número importante de situaciones (n=1174), siendo similar la cantidad de actividades analizadas por cada país. En Chile es muy superior la cantidad de actividades incluidas en los textos analizados en 7º curso (69,2%) que en 8º (30,8%), lo que puede deberse a que el estudio explícito de la tabla está propuesto en 7º básico, mientras que en el siguiente curso es un medio para la enseñanza de otros conceptos. En España, al contrario, el mayor número de situaciones se encuentran en 2º curso (56,3%) que en 1º curso de ESO (43,7%).

Tabla 1. Frecuencia (F) y porcentaje (%) de actividades analizadas, según nivel educativo y país
Fuente: Elaboración propia.

Nivel educativo	**Textos chilenos**		**Textos españoles**		**Total**	
	F	**%**	**F**	**%**	**F**	**%**
7º EB/1º ESO	393	69,2	265	43,7	658	56
8º EB/2º ESO	175	30,8	341	56,3	516	44
Total	568	100	606	100	1174	100

Para el análisis de las situaciones problema presentadas en los libros de texto se utilizó un análisis de contenido (Díaz-Herrera, 2018), el que permite profundizar en las características de documentos escritos, en este caso el libro de texto. El primer paso consiste en la selección de las unidades de análisis: para ello se identifica, en el tema o unidad didáctica de estadística y probabilidad de cada texto que compone la muestra, los párrafos que contuvieran situaciones en donde se hiciera uso de la tabla estadística.

Seguidamente, se determina la variable de análisis, en este caso cada una de las actividades, y en ellas se analiza el nivel de lectura y contexto. De este modo, mediante un proceso cíclico e inductivo (Bisquerra, 2014), se clasifican las actividades en que se involucra cualquier tipo de tabla estadística en las categorías definidas, seleccionando ejemplos de cada una de ellas. Se realizan continuas revisiones de los textos por parte de los autores, discutiendo los desacuerdos de codificación, para llegar a un consenso asegurando así la fiabilidad del análisis realizado. Finalmente, se elaboran tablas para resumir los resultados obtenidos y facilitar la obtención de conclusiones.

4. Resultados

En el siguiente apartado se presentan los resultados obtenidos en el análisis del nivel de lectura requerido junto al contexto en que es planteada la tabla estadística, en la muestra de libros de texto chilenos y españoles.

4.1 Niveles de lectura

En primer lugar, se describen los resultados sobre el nivel de lectura encontrados en los libros de texto analizados, utilizando los niveles de lectura propuestos por Friel et al. (2001). Se comienza describiendo con detalle las categorías de análisis, incorporando ejemplos que clarifiquen cada categoría, siguiendo con los resultados en los textos chilenos y luego en los textos españoles, para finalmente describir la comparación entre ambos países.

L1. *Leer los datos.* En este nivel, la respuesta a la pregunta planteada únicamente implica una lectura literal de la información que es explícita en la tabla, por lo que no se requiere realizar cálculos o ninguna otra operación con los datos representados.

El ejemplo reproducido en la Figura 1 presenta una actividad seleccionada de un libro de texto chileno dirigido a 7º básico, en el que luego de completar la tabla de distribución de una variable cualitativa con frecuencias ordinarias (absolutas, relativas y porcentuales), el estudiante debe responder a una serie de cuestiones entre ellas para identificar la variable de estudio y su tipo. Este tipo de preguntas es importante, pues de acuerdo a la variable representada se podría, posteriormente, determinar la pertinencia de calcular medidas de centralización como la media, moda o mediana. Otro tipo de pregunta que se plantea es identificar la frecuencia porcentual de una determinada categoría de la variable, y explicar cómo se obtuvo la respuesta, donde basta con indicar la columna utilizada para ello. Ambas cuestiones señaladas requieren de una lectura literal de la tabla, por lo que se enmarcarían en el nivel más básico de lectura, de acuerdo a Friel et al. (2001).

Se introducen las diferentes piezas de un ajedrez en una caja y se extrae una al azar, se anota qué pieza es y se devuelve a la caja.

Este experimento aleatorio se ha repetido 40 veces y se han obtenido los siguientes resultados:

	Peón	Torre	Caballo	Alfil	Rey	Reina
Frecuencia absoluta, n_i	7	5	10	3	8	7

a. ¿Cuál es el espacio muestral del experimento?

b. ¿Cuál es la frecuencia absoluta de cada una de las piezas?

Figura 1. Ejemplo de actividad con nivel de lectura L1, leer los datos. Fuente: EGB5 (p. 221).

L2. *Leer dentro de los datos.* En este nivel, la actividad no solo implica la lectura literal, sino que también se precisa comparar datos representados en la tabla, o aplicar cálculos con ellos. Se requiere de este nivel, por ejemplo, para determinar las medidas de centralización y rango de una distribución. Un ejemplo se presenta en la Figura 2, que es una actividad resuelta recogida de un libro de texto español dirigido a 2º ESO. En ella se explica al estudiante los pasos necesarios para calcular la media aritmética mostrando, además, la fórmula de cálculo y utilizando simbolización, e incluso subíndices. Para seguir el proceso de cálculo, el estudiante ha de leer en la tabla las modalidades de la variable y su frecuencia (absoluta y absoluta acumulada) correspondiente, que debe identificar en la tabla. Además, necesita comprender dichos conceptos para realizar correctamente las operaciones que se hacen con sus valores, según las columnas respectivas en la tabla, lo cual corresponde al nivel de leer dentro de los datos, ya que se pide al estudiante obtener información que no está representada literalmente en la tabla. Resaltamos el hecho de que se utiliza la notación simbólica (subíndices) para representar el valor de la variable xi, que se particulariza como x1, x2, x3 en la fórmula; de este modo, el estudiante debe deducir que i indica el orden de los valores de la variable en la tabla y a su vez informa del valor de la variable correspondiente a dicho subíndice en dicha fórmula.

La variable *Número de intentos* es una variable cuantitativa.

x_i	f_i	F_i	$f_i \cdot x_i$
1	27	27	$1 \cdot 27 = 27$
2	65	$27 + 65 = 92$	$2 \cdot 65 = 130$
3	8	$92 + 8 = 100$	$3 \cdot 8 = 24$
Total	$27 + 65 + 8 = 100$		$27 + 130 + 24 = 181$

- **Media aritmética** $\bar{x} = \dfrac{f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3}{N} = \dfrac{181}{100} = 1{,}81$
- **Mediana** El valor central ocupará la posición: $\dfrac{N}{2} = 50$. Por las frecuencias absolutas acumuladas sabemos que los 27 primeros datos son 1 y a partir de ahí, hasta la posición 92 son 2. Los datos que ocupan los lugares 50 y 51 son 2. $\rightarrow Me = (2 + 2)/2 = 2$
- **Moda** El valor más repetido, 65 veces, es 2 $\rightarrow Mo = 2$
- **Rango** Máximo = 3, mínimo = 1 $\rightarrow R = 3 - 1 = 2$

Figura 2. Ejemplo de actividad con nivel de lectura L2, leer dentro de los datos.
Fuente: ESO3 (p. 283).

L3. *Leer más allá de los datos*. Se trata de actividades que implican una mayor capacidad de lectura de la tabla estadística, pues se pide inferir información no representada y que no es posible extraer mediante operaciones aritméticas. Un ejemplo sería interpolar o extrapolar un valor en una serie de datos ordenados a lo largo del tiempo. En la Figura 3 se reproduce una actividad de un libro de texto chileno que incluye una pregunta cuya respuesta requiere este nivel de lectura. Se debe conjeturar, basándose en los datos de la tabla, el valor de la frecuencia relativa de cada modalidad de la variable si se incrementan las extracciones de una urna con fichas numeradas de 1 a 10. Para responder la pregunta, no solo se deben leer los valores de la frecuencia que ha correspondido a cada ficha numerada en la tabla (nivel de lectura 1), y comprobar que el valor de cada una es, aproximadamente, la décima parte de las extracciones (nivel de lectura 2). Además, se debe inferir que esta similitud de frecuencias se conservará en nuevos lanzamientos, por lo que, al realizar los 100.000 lanzamientos, aproximadamente cada número saldrá 10.000 veces. Además, hay que comprender que los resultados deben tener cierta variabilidad. Por tanto, se requiere un nivel de lectura 3.

Completa la tabla en tu cuaderno considerando que se realizaron 5000 extracciones de una ficha desde una urna con fichas numeradas desde el 1 al 10. Luego, resuelve.

5000 extracciones										
N°	1	2	3	4	5	6	7	8	9	10
f	490	513	501	491	508	506	493	498	502	498
f_r										

c. Si se aumentan las extracciones a 100 000, ¿a qué valor debiese tender cada probabilidad frecuencial? ¿Por qué?

Figura 3. Ejemplo de actividad con nivel de lectura L3, leer más allá de los datos.
Fuente: EGB1 (p. 213).

L4. *Leer detrás de los datos.* Corresponde al nivel más avanzado, que implica no solo la lectura del gráfico o tabla, sino también ser capaz de realizar una valoración crítica de su contenido, de las fuentes de las que se ha extraído la información, o de la veracidad de las afirmaciones realizadas sobre su contenido (Shaughnessy et al., 1996). En la Figura 4 se reproduce un ejemplo de un libro de texto español, en el que se pide al estudiante tomar una decisión a partir de la información entregada en tres tablas de frecuencia. Para ello, se debe discutir si tres escolares ficticios han realizado la tarea descrita correctamente, aunque todos alegaran que lo han hecho. Para ello, los estudiantes deben relacionar la amplitud de los sectores de la ruleta dibujada en el enunciado, uno de los cuales ocupa la mitad del área y, en consecuencia, aproximadamente, el color rojo ha de obtenerse la mitad de las veces (180). De los otros dos sectores, uno ocupa dos sextos del área y el otro uno, por lo cual los valores aproximados de veces en que aparecen los colores azul y amarillo serán 120 y 60, de lo que se deduce que Adrián ha hecho trampa. Puesto que no es de esperar unos valores exactos, Carla también ha debido inventar sus datos. En consecuencia, además de leer las tablas y comparar los datos con los esperados al realizar el experimento, se debe observar de manera crítica los resultados presentados en cada muestra y argumentar la respuesta, lo que supone el nivel superior de lectura.

Un profesor deja a cada uno de sus alumnos una ruleta como la del dibujo y les pide, para casa, que hagan girar la flecha 360 veces y que anoten los resultados.
Estos son los deberes entregados por tres alumnos. Dos de ellos han hecho trampa. ¿Cuáles crees que son? Explica por qué.

	ADRIÁN
ROJO	124
AZUL	126
VERDE	110

	MANUELA
ROJO	193
AZUL	111
VERDE	56

	CARLA
ROJO	180
AZUL	120
VERDE	60

Figura 4. Ejemplo de actividad con nivel de lectura L4, leer detrás de los datos.
Fuente: ESO1 (p. 303).

4.1.1 Resultados en los libros de textos chilenos

La distribución de los diferentes niveles de lectura presentados en las actividades propuestas en los libros de texto chilenos dirigidos a 7º y 8º básico se resume en la Tabla 2. Globalmente, el nivel de lectura que aparece con mayor frecuencia es L2, leer entre los datos (75,4%), que supone la realización de comparaciones y cálculo a partir de los datos de la tabla. Sigue el nivel L1, leer los datos (16,4%), es decir, la lectura directa sin operaciones posteriores. Los niveles más sofisticados de lectura como L3, leer más allá de los datos (2,6%) y L4, leer detrás de los datos (5,6%), que requieren interpolación o extrapolación y lectura crítica de la tabla, aparecen en un menor porcentaje.

Tabla 2. Frecuencia (F) y porcentaje (%) de niveles de lectura por nivel educativo. Fuente: Elaboración propia.

Nivel de lectura	**Educación Básica**				**Total**	
	7º EB		**8º EB**			
	F	**%**	**F**	**%**	**F**	**%**
L1. Leer los datos	66	16,8	27	15,4	93	16,4
L2. Leer dentro de los datos	312	79,4	116	66,3	428	75,4
L3. Leer más allá de los datos	10	2,5	5	2,9	15	2,6
L4. Leer detrás de los datos	5	1,3	27	15,4	32	5,6
Total	393	100	175	100	568	100

Hay un ligero aumento de nivel al progresar el curso, pues en 7º básico el nivel que se presenta en mayor porcentaje es L2, *leer entre los datos* (79,4%), mientras los niveles superiores de lectura como L3, *leer más allá de los datos* (2,5%) y L4, *leer detrás de los datos* (1,3%), se presentan en menor frecuencia. Aunque en 8º básico también es más frecuente el nivel L2, *leer entre los datos* (66,3%), su frecuencia disminuye con respecto al curso anterior, mientras que, en este nivel educativo, la lectura crítica de la tabla, es decir, el nivel L4, *leer detrás de los datos* (15,4%), se presenta en un porcentaje mayor que en 7º básico, con igual porcentaje que el nivel L1, *leer los datos*.

En la Figura 5 se analizan los niveles de lectura en las tres editoriales revisadas en el trabajo. En la figura se aprecia que en las tres editoriales predominan las actividades de nivel de lectura L2, *leer entre los datos* (MINEDUC, 74,4%; Santillana, 77,7%; SM, 74,5%), mientras que el nivel más básico de lectura L1, *leer los datos*, se presenta en un mayor porcentaje en los textos entregados por el MINEDUC (21,9%), seguido de la editorial Santillana (14,5%). Como se ha dicho, los niveles más avanzados de lectura son poco utilizados, siendo la editorial SM la que incluye un mayor porcentaje de actividades en los niveles L3, *leer más allá de los datos* (5,1%) y L4, *leer detrás de los datos* (8,7%).

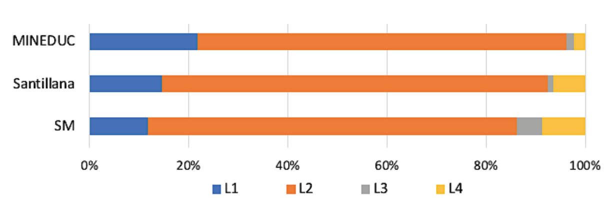


Figura 5. Porcentaje de actividades con diferentes niveles de lectura por editoriales chilenas.
Fuente: Elaboración propia.

4.1.2 Resultados en los libros de texto españoles

La Tabla 3 resume los resultados obtenidos en el análisis de los textos españoles respecto al nivel de lectura presentado en las actividades en los dos cursos analizados. Al igual como ocurrió en los textos chilenos, globalmente es más frecuente el segundo nivel, *leer entre los datos* (79%), seguido por el nivel inicial L1, *leer los datos* (19,8%). Los niveles superiores L3, *leer más allá de los datos* (0,2%) y L4, *leer detrás de los datos* (1%) son muy poco frecuentes.

Tabla 3. Frecuencia (F) y porcentaje (%) de niveles de lectura por nivel educativo. Fuente: Elaboración propia.

Nivel de lectura	**Educación Secundaria**				**Total**	
	1º ESO		**2º ESO**			
	F	**%**	**F**	**%**	**F**	**%**
L1. Leer los datos	48	18,1	72	21,1	120	19,8
L2. Leer dentro de los datos	215	81,1	264	77,4	479	79,0
L3. Leer más allá de los datos	1	0,4			1	0,2
L4. Leer detrás de los datos	1	0,4	5	1,5	6	1
Total	265	100	341	100	606	100

Al comparar la distribución de niveles de lectura en ambos cursos, sigue siendo el segundo nivel L2, *leer entre los datos*, el que aparece con mayor fuerza, aunque disminuye ligeramente en el segundo curso (1º ESO, 81,1%; 2º ESO, 77,4%). Le sigue el nivel inicial L1, *leer los datos*, que es algo más frecuente en 2º curso (21,1%) que en 1º ESO (18,1%). El nivel L3, *leer más allá de los datos*, aparece de manera muy escasa en 1º curso (0,4%) y no se presenta en 2º curso de ESO, mientras que L4, *leer detrás de los datos*, tiene una presencia algo más en 2º curso (1,5%) que en 1º curso ESO (0,4%).

La Figura 6 muestra la distribución de los niveles de lectura en las tres editoriales españolas analizadas. Observamos que las tres colecciones priorizan en las actividades el nivel de lectura L2, *leer entre los datos* (Anaya, 84,9%; Edelvives, 74,3%; Santillana, 77,7%), que prácticamente abarca la totalidad de las actividades, debido a que la lectura de tablas en los textos se usa como paso previo para el cálculo de estadísticos como la media, moda o el rango. En segundo lugar, de importancia se incluye el nivel L1, *leer los datos*; dicho nivel aparece menos en Anaya (14,6%), en comparación a las demás editoriales. Los niveles de lectura superiores apenas aparecen: notamos que L3, *leer más allá de los datos*, es recogido únicamente y de manera muy limitada por Edelvives (0,4%), esta última editorial es la que presenta un porcentaje ligeramente superior de actividades que implican el máximo nivel de lectura L4, *leer detrás de los datos* (2,2%), dicho nivel también es recogido por Anaya, pero de forma muy escasa (0,5%).

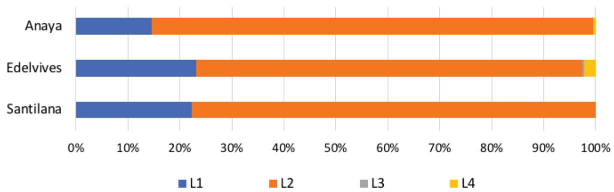


Figura 6. Porcentaje de niveles de lectura por nivel educativo y editoriales españolas. Fuente: Elaboración propia.

4.1.3 Comparación de resultados en los dos países

Para completar el análisis de las secciones anteriores, en la Figura 7 se presenta un gráfico comparativo con la distribución de los niveles de lectura por nivel educativo y país. Observamos que los libros de texto chilenos presentan una mayor proporción de actividades de niveles L3, *leer más allá de los datos* y L4, *leer detrás de los datos* que los españoles. También se aprecia un incremento de los niveles de lectura conforme se progresa de curso, pues en 7º básico los niveles L3, *leer más allá de los datos* (2,5%) y L4, *leer detrás de los datos* (1,4%) aparecen de manera muy escasa, pero en 8º se incrementan de manera significativa (L3, 2,9%; L4, 15,4%).

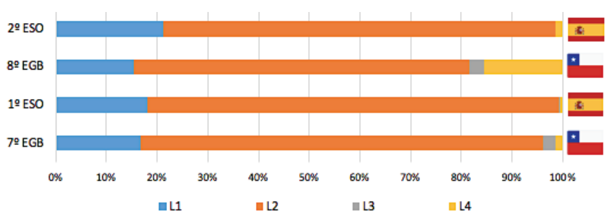


Figura 7. Porcentaje de actividades con diferentes niveles de lectura por nivel educativo y país. Fuente: Elaboración propia.

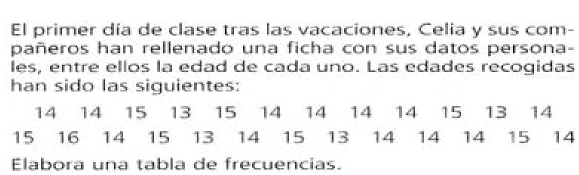
Los libros de texto españoles también muestran un incremento del nivel máximo de lectura L4, *leer detrás de los datos* (1º ESO, 0,4%; 2º ESO, 1,5%), aunque de manera más ligera, conforme se avanza de nivel educativo; sin embargo, L3, *leer más allá de los datos* no es contemplado en 2º curso. En estos textos se prioriza actividades que requieren de los niveles de lectura literal, por ejemplo, L1, *leer los datos* se incrementa de 1º curso (18,1%) a 2º curso ESO (21,1%), mientras que L2, *leer entre los datos* disminuye ligeramente (1º ESO, 81,1%; 2º ESO, 77,4%).

En resumen, los resultados obtenidos del análisis muestran una predominancia de actividades que requieren los niveles más básicos de lectura, L1, leer los datos y L2, *leer entre los datos* en los dos países, que coinciden con los resultados obtenidos por Díaz-Levicoy et al. (2015, 2018) con libros de texto de los primeros años de educación primaria en Chile. En nuestro caso es mayor la proporción del nivel L2, *leer entre los datos*, y menor del nivel L1, *leer los datos* respecto a los datos obtenidos por dichos autores, lo que se comprende al tratarse de cursos superiores a los que ellos analizaron. Lo mismo ocurrió en la investigación de García-García et al. (2019) en México, y en todas estas investigaciones los niveles L3, *leer más allá de los datos* y L4, *leer detrás de los datos* no se consideraron, también razonable porque dichos trabajos se dirigen a niveles educativos inferiores a los tratados en, por ejemplo, nuestro caso. En nuestro análisis anterior de otros textos chilenos (Pallauta et al., 2021) se obtuvieron resultados muy similares en cuanto a la alta presencia de los niveles más básicos de lectura, mientras en el presente estudio se observa un incremento de los niveles más sofisticados de lectura, especialmente en 8º básico.

4.2 Contextos

Seguidamente se presentan los resultados del análisis de los contextos en que se plantean los problemas en los textos de la muestra. Comenzamos describiendo las categorías de contextos, que corresponden a los planteados en los estudios PISA (OECD, 2019), a las que se ha añadido "experimento aleatorio" como un nuevo contexto, debido a su uso en los textos analizados para el estudio de la probabilidad frecuencial. Luego se presenta la distribución de contextos en los textos chilenos y españoles, para finalizar con una comparación entre ambas colecciones de textos.

Contexto personal. Entre las situaciones consideradas en las pruebas PISA, las de tipo personal se refieren a las cercanas a la experiencia inmediata del estudiante (Turner, 2006). Son las actividades propias del cotidiano, o las que puede observar en su entorno inmediato. Por ejemplo, en la Figura 8 se presenta la distribución de los deportes favoritos de un grupo de estudiantes. Otros contextos personales encontrados en los textos son las actividades que prefieren los estudiantes realizar en su tiempo libre, o la cantidad de goles anotados por su equipo favorito en un partido de fútbol.


El primer día de clase tras las vacaciones, Celia y sus compañeros han rellenado una ficha con sus datos personales, entre ellos la edad de cada uno. Las edades recogidas han sido las siguientes:

14 14 15 13 15 14 14 14 14 15 13 14
15 16 14 15 13 14 15 13 14 14 14 15 14

Elabora una tabla de frecuencias.

Figura 8. Ejemplo de actividad de contexto personal. Fuente: ESO5 (p. 200).

Contexto laboral. Son situaciones algo más apartadas de lo personal, que cualquiera podría encontrar en una actividad diaria organizada, como el trabajo (Turner, 2006). Es decir, las actividades propias del quehacer ocupacional o profesional que se desarrollan en diversos ámbitos. La Figura 9 es una actividad seleccionada de un libro de texto español, dirigida a 2º curso de secundaria, y muestra la cantidad de libros vendidos por los tres dependientes de una librería. A partir de esta información se pide determinar algunos resúmenes estadísticos.

En una librería, el número de libros que han vendido tres trabajadores en una semana de trabajo ha sido el siguiente:

	L	M	X	J	V	S
Anna	4	2	7	3	8	3
Quim	5	4	6	6	4	3
Clara	4	4	4	4	4	4

a) Calcula las medidas estadísticas de lo que ha vendido cada trabajador.

Figura 9. Ejemplo de actividad de contexto ocupacional o laboral.
Fuente: ESO6 (p. 291).

Contexto público. Se trata de las situaciones presentes en la comunidad y el entorno del estudiante, que se pueden observar en el entorno local o en contextos más amplios, a las que es posible de acceder a través de los diferentes medios de comunicación. En la Figura 10 se muestra un ejemplo recogido de un libro de texto chileno de 7º básico en que se exponen los resultados sobre una consulta, realizada por un organismo gubernamental, en que se les pide a los jóvenes caracterizar a las personas que viven en situación de pobreza. Posteriormente, con dicha información, el estudiante debe construir una tabla de distribución.

El 41 % de los jóvenes opinan que la mayoría de las personas en situación de pobreza en Chile son personas que no alcanzan condiciones de vida dignas en salud, trabajo, educación y vivienda. El 23 % de los entrevistados opina que son personas que viven en hogares que no consiguen obtener un mínimo de ingreso al mes, mientras que el 17 % opina que son personas que son discriminadas o excluidas por su origen social, cultura, género u otras razones. Es interesante observar que otro 17 % de los entrevistados opina que son personas que no trabajan o no se esfuerzan. Un 2 % no sabe o no responde.

Información extraída de: http://www.injuv.gob.cl/storage/docs/Sondeo_Percepcion_de_los_jovenes_sobre_la_pobreza.pdf

Si los casos fueron 1010, construye una tabla de frecuencias que represente los datos mencionados.

Figura 10. Ejemplo de actividad de contexto público.
Fuente: EGB1 (p. 194).

Contexto científico. Incluyen aplicaciones de las matemáticas más abstractas y complejas que los anteriores contextos, y se asocian a procesos científicos o tecnológicos. Algunos ejemplos serían la duración de la inmunización tras recibir alguna de las dosis de la vacuna para el COVID-19, o la relación del grupo sanguíneo y el factor RH de un grupo de personas. En la Figura 11 se muestra un ejemplo de un libro de texto chileno dirigido a 8º básico en que se presenta una tabla de datos con el registro de las temperaturas alcanzadas en el océano Pacífico según la latitud.

Los rayos del sol caen de forma más directa en las regiones cercanas a la línea del ecuador, durante casi todo el año.

La tabla muestra el promedio de las temperaturas registradas en el océano Pacífico según sea la latitud.

Promedios de temperaturas	
Latitud	Temperatura (ºC)
0º, 10º	26,0
20º, 30º	21,5
50º, 60º	5,0
60º, 70º	–1,3

Figura 11. Ejemplo de actividad de contexto científico.
Fuente: EGB6 (p. 266).

Experimento aleatorio. Al igual que Pallauta et al. (2021), incorporamos esta categoría al análisis del contexto debido a su alta presencia en actividades en que se presentan los resultados en juegos de azar o experiencias aleatorias, que son planteadas para el estudio de la probabilidad, principalmente, desde su enfoque frecuencial (Ortiz, 2002). Las Figuras 1, 3 y 4 son ejemplos de ello.

Sin contexto. Consideramos en esta última categoría las actividades en que no se presenta el contexto en que fueron extraídos los datos presentados en la tabla estadística, como en la Figura 12, que muestra un ejemplo de un libro de texto español de 2º ESO. En este sentido, como señala Van den Heuvel-Panhuizen (2005), se debe tener en cuenta que una enseñanza desvinculada de las experiencias de los estudiantes podría significar que olviden de forma rápida lo aprendido, mientras que un aprendizaje producido mediante tareas en contexto promueve mejores opciones de comprensión de conceptos a largo plazo.

2 Copia esta tabla de frecuencias y complétala.

Datos	Frecuencia absoluta	Frecuencia relativa
2		0,25
3	7	
7		0,125
20	13	
	$N = 32$	

Figura 12. Ejemplo de actividad sin contexto.
Fuente: ESO6 (p. 293).

4.2.2 Análisis de contextos en los textos chilenos

Los diferentes tipos de contextos de las actividades planteadas respecto a las tablas estadísticas en los libros de texto chilenos son resumidos en la Tabla 4. Apreciamos que el contexto que aparece con mayor fuerza es el de tipo personal (46,7%), que muestra el interés de los autores de los textos por acercar la estadística a los intereses inmediatos de los estudiantes. Sigue en frecuencia el contexto de experimento aleatorio (19,9%) requerido en el trabajo con la probabilidad, y a continuación el público (15,1%), relacionado con la comunidad del estudiante.

Tabla 4. Frecuencia (F) y porcentaje (%) de tipo de contexto por nivel educativo. Fuente: Elaboración propia.

Contexto	**Educación Básica**				**Total**	
	7º EB		**8º EB**			
	F	**%**	**F**	**%**	**F**	**%**
Personal	185	47,1	80	45,7	265	46,7
Laboral/ educativo	19	4,8	31	17,7	50	8,8
Público	67	17	19	10,9	86	15,1
Científico	16	4,1	22	12,6	38	6,7
Experimento aleatorio	93	23,7	20	11,4	113	19,9
Sin contexto	13	3,3	3	1,7	16	2,8
Total	393	100	175	100	568	100

Al comparar el contexto en los dos cursos analizados, se puede apreciar que en 7º básico se presenta más el contexto personal (47,1%) que en 8º curso (45,7%), igual que ocurre con el experimento aleatorio (7º curso, 23,7%; 8º básico, 11,4%). En 8º curso observamos que el contexto ocupacional aparece de manera significativa (17,7%), a diferencia de 7º básico (4,8%) y lo mismo ocurre con el contexto científico (12,6% en 8º frente a 4,1% en 7º). Pensamos que los autores de los textos han querido incidir más en estos temas más abstractos con los estudiantes de mayor edad. Las actividades sin contexto son poco utilizadas tanto en 7º (3,3%) como en 8º curso (1,7%).

Los diferentes tipos de contextos, según editorial, son presentados en la Figura 13, en la que se puede notar que las editoriales priorizan el uso del contexto personal para plantear las tablas, siendo más frecuente en Santillana (52,2%). El contexto laboral se presenta de manera similar en Santillana (13,4%) y SM (11,7%), mientras que MINEDUC lo recoge en menor medida (2,8%), y en cambio le da mayor relevancia al contexto público, a diferencia de las demás editoriales. La categoría experimento aleatorio aparece de manera significativa en MINEDUC (24,2%) y SM (25,5%), y en Santillana tiene una menor frecuencia (7%). El contexto científico no es incorporado por MINEDUC, a diferencia de Santillana que lo incluye en una proporción apreciable (18,5%). Destacamos que las actividades sin contexto son muy escasas en MINEDUC (0,9%) y Santillana (0,6%), y se presentan algo más en SM (6,6%).

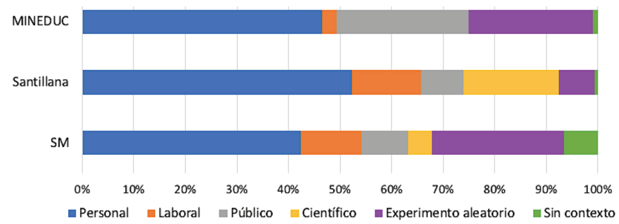


Figura 13. Porcentaje de contextos, según editorial chilena. Fuente: Elaboración propia.

4.2.3 Análisis de contextos en los textos españoles

La Tabla 5 resume los tipos de contextos encontrados en los libros de texto españoles. Al igual que en los textos chilenos, el contexto con mayor presencia es el personal (48,8%). En el caso de los textos españoles el segundo más frecuente es el contexto público (15,7%), seguido por el experimento aleatorio (11,4%).

Al comparar los dos cursos de la muestra de textos, el contexto personal aparece con mayor fuerza en 1º curso (57,7%) que en 2º curso de la ESO (41,9%); lo mismo ocurre con las actividades sin contexto que son más frecuentes en 1º (11,3%) que en 2º curso (8,8%), lo cual va en contra de los principios didácticos, puesto que dichas actividades son más abstractas. El contexto experimento aleatorio también en estos textos tiene una presencia significativa, siendo más utilizado en 2º curso (12,6%) que en 1º curso de la ESO (9,8%).

Tabla 5. Frecuencia (F) y porcentaje (%) de tipo de contexto por nivel educativo. Fuente: Elaboración propia.

Contexto	**Educación Secundaria**				**Total**	
	1º ESO		**2º ESO**			
	F	**%**	**F**	**%**	**F**	**%**
Personal	153	57,7	143	41,9	296	48,8
Laboral/ educativo	11	4,2	41	12	52	8,6
Público	38	14,3	57	16,7	95	15,7
Científico	7	2,6	27	7,9	34	5,6
Experimento aleatorio	26	9,8	43	12,6	69	11,4
Sin contexto	30	11,3	30	8,8	60	9,9
Total	265	100	341	100	606	100

Al analizar la presencia de los diferentes tipos de contextos según editorial en los textos españoles, observamos en la Figura 14 que, al igual que las editoriales chilenas, se prioriza el uso del contexto personal, siendo Edelvives la que menos lo utiliza (37%) respecto a las demás. El contexto laboral se presenta de manera similar en Santillana (14,6%) y Edelvives (11,3%), pero es muy escaso en Anaya (1,4%). Por otra parte, experimento aleatorio aparece en todas las editoriales, pero con diferente peso, siendo más frecuente en Edelvives (20,4%), seguido de Anaya (8,7%) y con una escasa presencia en Santillana (1,9%). Las actividades sin contexto son pocas en Edelvives (3%), mientras que en Anaya (14,6%) y Santillana (13,4%) se presentan en un mayor porcentaje.

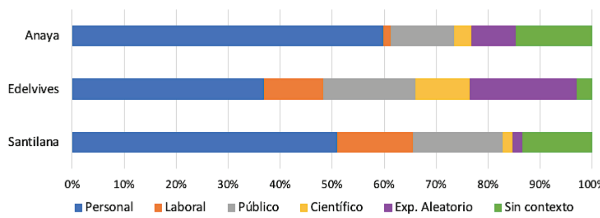


Figura 14. Porcentaje de contextos, según editorial española.
Fuente: Elaboración propia.

4.2.4 Comparación de resultados en los dos países

En la Figura 15 se presenta un gráfico comparativo que muestra la distribución de los tipos de contextos, según nivel educativo y país. Apreciamos que en ambos países se prioriza el uso del contexto de tipo personal, lo que se enmarca en los lineamientos curriculares (MECD, 2015; MINEDUC, 2015), en cuanto a la importancia de proponer situaciones que sean cercanas para el estudiante. Este resultado coincide con otros análisis del contexto en que se proponen las tablas estadísticas en los libros de texto dirigidos a diferentes niveles educativos (Díaz-Levicoy et al., 2015, 2018; García-García et al., 2019; Pallauta et al., 2021).

En los textos chilenos se aprecia un incremento del contexto laboral conforme se progresa de curso y, de manera contraria, se disminuye la presencia de las actividades sin contexto al avanzar de 7º a 8º básico. En los libros de texto españoles se aprecia un incremento de los contextos laboral, científico y experimento aleatorio conforme se avanza de 1º a 2º ESO. Notamos que las actividades sin contexto tienen una mayor presencia en los libros españoles, a diferencia de los chilenos en que son menos utilizadas.

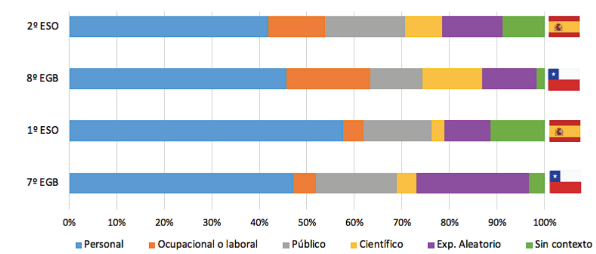


Figura 15. Porcentaje de contextos por nivel educativo y país.
Fuente: Elaboración propia.

5. Conclusiones

El análisis presentado proporciona nueva información referente al nivel de lectura requerido y el contexto en que se presentan las actividades sobre tablas estadísticas, en una muestra de libros de texto chilenos y españoles, siendo el primer estudio que se realiza sobre este tema en textos españoles, y complementando otro anterior (Pallauta et al., 2021) llevado a cabo con textos chilenos. Ambas variables han de ser tenidas en cuenta en la enseñanza, pues los niveles de lectura, de los propuestos por Curcio y colaboradores (Curcio, 1989; Friel et al., 2001), permiten identificar la demanda cognitiva requerida de las tareas propuestas, mientras que el contexto posee un rol relevante en el desarrollo del razonamiento estadístico de los estudiantes (Makar y Ben-Zvi, 2011).

Aunque los resultados sobre los niveles de lectura muestran un ligero incremento de los niveles de lectura más sofisticados, especialmente *leer detrás de los datos,* conforme se progresa de curso, especialmente en la muestra de textos chilenos, llama la atención la gran cantidad de actividades que se centran en los niveles más básicos de lectura. Aunque estos resultados coinciden con otros estudios, estos abordan niveles educativos inferiores (Díaz-Levicoy et al., 2015; García-García et al., 2019). Será importante entonces que, conforme se avanza de curso, se incremente el nivel de dificultad de las tareas, sobre todo fomentando una lectura crítica de los datos, necesaria como parte de la cultura estadística para tomar mejores decisiones, de cara al escenario de incertidumbre que enfrenta la sociedad en su conjunto (Rodríguez-Muñiz et al., 2020).

En relación a los contextos PISA (OECD, 2019), destacamos la alta presencia del contexto de tipo personal, tanto en los textos chilenos como españoles, que también concuerda con estudios anteriores (Díaz-Levicoy et al., 2015, 2018; García-García et al., 2019; Pallauta et al., 2021), y es sugerido en los lineamientos curriculares de ambos países (MECD, 2015; MINEDUC, 2015) en que se destaca la importancia de plantear contextos cercanos para el estudiante. Resaltamos también la frecuencia de uso del contexto experimento aleatorio, el cual fue incluido, al igual que en Pallauta et al. (2021), en el estudio de la probabilidad, pues las experiencias de juegos de azar forman parte del entorno cercano, y pueden permitir al profesor diseñar una enseñanza de la estadística basada en datos reales (Alsina y Annexa, 2021) que pueda ser más significativa para el estudiante. En este sentido, llama la atención la presencia de actividades que carecen de contexto, que, a pesar de ser escasas, especialmente en los textos chilenos, alerta de una tendencia que se pudiera incrementar en los cursos superiores, pues este tipo de actividades no facilita el equilibrio que debiera existir entre el contexto de una situación y la estadística que se aplica para resolverla (Wild y Pfannkuch, 1999).

La información presentada puede ser de interés para profesores, quienes la deben tener en cuenta en el diseño de la enseñanza de este tema, en particular. Puesto que la investigación sobre temas de estadística en educación secundaria, en general, es escasa, también se aporta información para la investigación en educación estadística, al realizar un comparativo de dos países sobre el tratamiento de la tabla estadística en los libros de texto. Como señalamos anteriormente, el libro de texto es ampliamente utilizado en el aula (Herbel, 2007), y en la actualidad toma mayor relevancia, pues en muchos casos es el único recurso con que cuentan los estudiantes, debido a que, en diferentes regiones y países, los procesos de instrucción no se están pudiendo realizar de manera presencial producto de la pandemia generada por el COVID-19.

Agradecimientos

Proyecto PID2019-105601GB-I00 / AEI / 10.13039/501100011033, Grupo FQM126 (Junta de Andalucía) y Beca ANID Folio: 72190280.

Referencias

Alsina, Á., y Annexa, E. (2021). Estadística en contexto: Desarrollando un enfoque escolar común para promover la alfabetización. *TANGRAM, 4*(1), 71-98. https://doi.org/10.30612/tangram.v4i1.14396

Bisquerra, R. (2014). *Metodología de la investigación educativa* (6.a ed.). La Muralla.

Blum, W., Galbraith, P., Henn, H.-W., y Niss, M. (Eds.). (2007). *Modelling and applications in mathematics education*. Springer. https://doi.org/10.1007/978-0-387-29822-1

Corral, Y., Corral, I., y Franco Corral, A. (2015). Procedimientos de muestreo. *Revista Ciencias de la Educación, 26*(46), 10-39.

Curcio, F. R. (1989). *Developing graph comprehension. Elementary and middle school activities.* National Council of Teachers of Mathematics.

Díaz-Herrera, C. (2018). Investigación cualitativa y análisis de contenido temático. Orientación intelectual de revista Universum. *Revista General de Información y Documentación, 28*(1), 119-142. https://doi.org/10.5209/RGID.60813

Díaz-Levicoy, D., Morales, R., y López Martín, M. (2015). Tablas estadísticas en libros de texto chilenos de 1o y 2o año de educación primaria. *Revista Paranaense de Educação Matemática, 4*, 10-39. https://doi.org/10.30612/tangram.v1i2.7574

Díaz-Levicoy, D., Vásquez, C., y Molina-Portillo, E. (2018). Estudio exploratorio sobre tablas estadísticas en libros de texto de tercer año de educación primaria. *Tangram, 1*(2), 18-39. https://doi.org/10.30612/tangram.v1i2.7574

Estrella, S., y Estrella, P. (2020). Representaciones de datos en estadística: De listas a tablas. *Revista Chilena de Educación Matemática, 12*(1), 21-34. https://doi.org/10.46219/rechiem.v12i1.20

Font, V. (2007). Comprensión y contexto: Una mirada desde la didáctica de las matemáticas. *Gaceta de la Real Sociedad Matemática Española, 10*(2), 427-442.

Friel, S. N., Curcio, F. R., y Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124-158. https://doi.org/10.2307/749671

Gal, I. (2019). Understanding statistical literacy: About knowledge of contexts and models. En J. M. Contreras, M. M. Gea, M. M. López-Martín, y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística.* https://www.ugr.es/~fqm126/civeest/ponencias/gal.pdf

García-García, J., Díaz-Levicoy, D., Vidal-Henry, S., y Arredondo, E. H. (2019). Las tablas estadísticas en libros de texto de educación primaria en México. *Paradigma*, 153-175. https://doi.org/10.37618/PARADIGMA.1011-2251.2019.p153-175.id754

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal, 16*(1), 22-25. https://doi.org/10.52041/serj.v16i1.209

Herbel, B. A. (2007). From Intended Curriculum to Written Curriculum: Examining the "Voice" of a Mathematics Textbook. *Journal for Research in Mathematics Education, 38*(4), 344-369. https://doi.org/10.2307/30034878

Makar, K., y Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1-2), 1-4. https://doi.org/10.1080/10986065.2011.538291

Ministerio de Educación, Cultura y Deporte de España. (2014). *Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria.* Ministerio de Educación, Cultura y Deporte, Boletín Oficial del Estado Núm. 52.

Ministerio de Educación, Cultura y Deporte de España. (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.* Ministerio de Educación, Cultura y Deporte, Boletín Oficial del Estado Núm. 3.

Ministerio de Educación de Chile. (2015). *Bases curriculares 7o básico a 2° medio.* Unidad de Currículum y Evaluación.

Ministerio de Educación de Chile. (2018). *Bases curriculares Primero a Sexto Básico.* Unidad de Currículum y Evaluación.

Organización para la Cooperación y el Desarrollo Económicos. (2019). PISA 2018 Mathematics Framework. En *PISA 2018 Assessment and Analytical Framework* (pp. 73-95). OECD. https://doi.org/10.1787/13c8a22c-en

Ortiz, J. J. (2002). *La probabilidad en los libros de texto.* Grupo FQM126.

Pallauta, J., y Arteaga, P. (2021). Niveles de complejidad semiótica en gráficos y tablas estadísticas. *Números, 106*, 13-22.

Pallauta, J., Gea, M. M., y Arteaga, P. (2021). Caracterización de las tareas propuestas sobre tablas estadísticas en libros de texto chilenos de educación básica. *Paradigma, 42*, 32-60. https://doi.org/10.37618/PARADIGMA.1011-2251.2021.p32-60.id1017

Pepin, B., y Gueudet, G. (2020). Curriculum resources and textbooks in mathematics education. En *Encyclopedia of mathematics education* (pp. 87-94). Springer. https://doi.org/10.1007/978-3-030-15789-0_40

Reys, B., Reys, R., y Chavez, O. (2004). Why mathematics textbooks matter. *Educational Leadership, 61*(5), 61-66.

Rodríguez-Muñiz, L. J., Muñiz-Rodríguez, L., Vásquez, C., y Alsina, Á. (2020). ¿Cómo promover la alfabetización estadística y de datos en contexto? Estrategias y recursos a partir de la COVID-19 para Educación Secundaria. *Números, 104*, 217-238.

Rosales, E. M., Rodríguez, P. G., y Romero, M. (2020). Conocimiento, demanda cognitiva y contextos en la evaluación de la alfabetización científica en PISA. *Eureka, 17*(2), 2302. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2020.v17.i2.2302.

Sanmartí, N., y Márquez, C. (2017). Aprendizaje de las ciencias basado en proyectos: Del contexto a la acción. *Ápice. 1*(1), 3-16. https://doi.org/10.17979/arec.2017.1.1.2020.

Shaughnessy, J. M., Garfield, J., y Greer, B. (1996). Data handling. En A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, y C. Laborde (Eds.), *International handbook of mathematics education* (Vol. 4, pp. 205-237). Springer. https://doi.org/10.1007/978-94-009-1465-0_8

Turner, R. (2006). El Programa Internacional para la Evaluación de los Alumnos (PISA). Una perspectiva general. *Revista de Educación*, nº extraordinario, 45-74.

Van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics, 25*(2), 2-23.

Wijaya, A., Van den Heuvel-Panhuizen, M., Doorman, M., y Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *The Mathematics Enthusiast, 11*(3), 555-584.

Wild, C. J., y Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review, 67*(3), 223-248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Apéndice: Muestra de libros utilizados

Textos chilenos

Código	**Referencia**
EGB1	Iturra, F., Manosalva, C., Ramírez, M., y Romero, D. (2019). 7º básico Matemática Texto del estudiante. SM. Edición especial para el Ministerio de Educación.
EGB2	Marambio, V., y Castro, C. (2016). Matemática 7º básico, Proyecto Todos Juntos. Santillana.
EGB3	Schwerter, S., Aguilar, M., y Maulén, M. (2014). Sé protagonista, Matemática 7º. SM.
EGB4	Torres, C. y Caroca, V. (2019). 8º Básico Matemática Texto del estudiante. Santillana. Edición especial para el Ministerio de Educación.
EGB5	Castro, C., Curiche, A., y Vega, M. (2014). Sé protagonista, Matemática 8. SM.
EGB6	Ramírez, A., Maldonado, L., Castro, C., y Ávila, J. (2016). Matemática 8º básico. Santillana.

Textos españoles

Código	**Referencia**
ESO1	Colera, J., Gaztelu, I., y Colera R. (2016). ESO 1 Matemáticas (Proyecto Aprender es crecer en conexión). Anaya.
ESO2	Mejía D., Romero, R., y Ocaña, J. (2015). ESO 1 Matemáticas (Proyecto somoslink). Edelvives.
ESO3	Almodóvar, J., De la Prida, C., Gaztelu, A., González, A., Machín, P., Pérez, C., y Sánchez, D. (2016). Matemáticas Serie Resuelve ESO 1 (Proyecto Saber Hacer). Santillana.
ESO4	Colera, J., Gaztelu, I., y Colera R. (2017). ESO 2 Matemáticas (Proyecto Aprender es crecer en conexión). Anaya.
ESO5	Romero, R., Ocaña, J., y Mejía D., (2016). ESO 2 Matemáticas (Proyecto somoslink). Edelvives.
ESO6	Almodóvar, J., Cuadrado A., Díaz, L., Dorce, C., Gámez, J., Marín, S., y Sánchez, D. (2016). Matemáticas Serie Resuelve ESO 2 (Proyecto Saber Hacer). Santillana.

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

ARTÍCULOS DE INVESTIGACIÓN

https://doi.org/10.46219/rechiem.v13i4.81

CONOCIMIENTO ESTADÍSTICO ESPECIALIZADO EN PROFESORES DE EDUCACIÓN BÁSICA, BASADO EN LA TAXONOMÍA SOLO

SPECIALIZED STATISTICAL KNOWLEDGE IN PRIMARY EDUCATION TEACHERS, BASED ON THE SOLO TAXONOMY

Pedro Vidal-Szabó
pedro.vidal@pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Soledad Estrella
soledad.estrella@pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

RESUMEN

El conocimiento especializado de los profesores que enseñan estadística en educación básica requiere de un modelo específico en educación estadística. Esta investigación explora en los conocimientos y niveles de comprensión que manifiestan 192 docentes sobre los conceptos de variable estadística, dato, información estadística y contexto de los datos como temas del eje curricular *datos y probabilidades* de la asignatura Matemática, lo que permite precisar una propuesta de extensión del modelo de Conocimiento Especializado del Profesor de Matemática (MTSK) a la Estadística, rotulado como modelo STSK (*i. e., Statistics Teacher´s Specialised Knowledge*). Por medio de la taxonomía SOLO se clasificaron 768 respuestas escritas dadas a cuatro ítems de un cuestionario en línea, instrumento que fue validado a través de juicio experto por 15 especialistas. Este estudio permite aportar teóricamente al conocimiento sobre la formación docente, pues se caracterizaron ciertos conocimientos estadísticos especializados del profesorado, a través de categorías y descripciones, siendo todavía un desafío la comprensión de variable estadística, dato, información estadística y contexto de los datos como conceptos estadísticos fundamentales para el desarrollo profesional docente.

PALABRAS CLAVE:
Estadística Temprana, Modelo STSK, Conocimiento Profesional Docente, Sentido del Dato.

ABSTRACT

The specialized knowledge of teachers who teach statistics in primary education requires a specific model in statistics education. This research explores the knowledge and levels of understanding declared by 192 teachers about the concepts of statistical variable, data, statistical information and context of data as topics of the curricular axis Data and Probabilities of the mathematics subject, which allows specifying an extension proposal of the Mathematics Teacher's Specialized Knowledge (MTSK) model to Statistics, labeled as STSK model (i.e., Statistics Teacher's Specialized Knowledge). The SOLO taxonomy was used to classify 768 written responses to four items of an online questionnaire, an instrument that was validated through expert judgment by 15 specialists. This study allows a theoretical contribution to the knowledge on teacher education, since certain specialized statistical knowledge of teachers was characterized through categories and descriptions, being still a challenge to understand the concepts of statistical variable, data, statistical information and context of data as fundamental statistical concepts for the professional development of teachers.

KEYWORDS:
Early Statistics, STSK Model, Professional Teaching Knowledge, Data Sense.

Recibido: 30 de Mayo de 2021, Aceptado: 6 de Agosto de 2021

1. Introducción

La estadística escolar necesita que los profesores comprendan que existe una manera diferente de razonar con datos reales, pues estos son números con un contexto que provee significado e integra la incertidumbre (Cobb y Moore, 1997; Estrella, 2018). La disciplina estadística es autónoma y no es una rama de la disciplina matemática (Wild et al., 2018), aunque los contenidos estadísticos a nivel curricular en Chile se ubiquen en un eje temático dentro de la asignatura matemática (Ministerio de Educación de Chile, 2012). Hace más de dos décadas, se conoce un marco para caracterizar el pensamiento estadístico que una persona activa durante una investigación empírica, cuya repercusión perdura en la comunidad de estadísticos educativos, pues distingue que el pensar estadístico tiene una naturaleza distinta al pensar matemático (Wild y Pfannkuch, 1999). Algunos estudios indican que el trabajo estadístico involucra a la matemática y la computación como herramientas auxiliares durante la modelación estadística (Pfannkuch, 2011; Wild et al., 2018).

Para promover el desarrollo del pensamiento estadístico en la etapa escolar, es preciso satisfacer algunos desafíos en la enseñanza. Por ejemplo, crear conciencia de las características del pensamiento estadístico activado en una variedad de contextos para, posteriormente, desarrollar estrategias de enseñanza que fomenten y promuevan este pensar en los estudiantes, dándoles acceso a convertir datos en visiones del mundo real (Wild et al., 2018). Para ello, la integración de la estadística con el contexto es fundamental porque ayuda a los estudiantes a comprender que la estadística no se desarrolla lejos de los problemas reales (Del Pino y Estrella, 2012). En consecuencia, el contexto en estadística implica al menos dos perspectivas: en lo epistemológico, el contexto remite a la situación del mundo real de la que emergen los datos; y en lo cognitivo, el contexto del aprendizaje y la experiencia dan cuenta de los conocimientos contextuales de la persona en un rol estadístico frente a hechos de interés (Vidal-Szabó et al., 2020).

En la sociedad de la información, es esperable que las personas lleguen a desarrollar cierto sentido del dato para el buen ejercicio de su ciudadanía. Estrella (2018) releva los datos y propone una caracterización sobre el sentido de los datos en relación con el desarrollo del aprendizaje de la estadística escolar, implicando dar solución a un problema estadístico con un cierto sentido numérico en contexto (Estrella et al., 2020). El sentido del dato involucra conocimientos, intuiciones y habilidades que una persona desarrolla acerca de los datos, especialmente en su escolaridad, permitiéndole flexibilidad y creatividad al resolver problemas estadísticos, utilizando argumentos y estrategias estadísticas propias (Estrella et al., en prensa). En particular, el sentido del dato involucra que un sujeto desarrolle un conjunto de conocimientos (e. g., representar datos de múltiples maneras según el contexto), intuiciones (e. g., expresarse respecto a los datos sin realizar ningún cálculo) y habilidades (e. g., hacer comparaciones entre diferentes representaciones de datos).

Algunos estudios han precisado algunas ideas fundamentales como datos, distribución, variabilidad, representación, asociación y modelación de relaciones entre variables, modelos de probabilidad, muestreo e inferencia (Burrill y Biehler, 2011; Garfield y Ben-Zvi, 2008). Dichas ideas requieren ser comprendidas a un nivel conceptual profundo por el profesorado encargado de su enseñanza de modo que, al menos, permita alfabetizar estadísticamente a la ciudadanía (Estrella, 2017).

En la etapa escolar, los docentes pueden desarrollar un sentido del dato que integra de modo funcional tanto los conceptos como los procedimientos estadísticos, por ejemplo, instando a sus estudiantes a visualizar las representaciones de datos como un todo imbricado que permite tomar conciencia del comportamiento de los datos, transitando desde la observación individual de datos puntuales al conjunto de los datos como agregados (Estrella, 2018; Estrella et al., 2020; Konold et al., 2015). Dado que los profesores de educación básica inician la alfabetización estadística a través de actividades que permiten desarrollar el sentido del dato, es necesario que la formación docente promueva el trabajo estadístico con datos reales de variables cualitativas y cuantitativas en contextos motivadores y de interés que favorezcan el aprendizaje estadístico. La presente investigación busca responder: ¿qué conocimientos y niveles de comprensión expresan profesores chilenos que enseñan estadística en educación básica sobre los conceptos de variable estadística, dato, información estadística y contexto de los datos? Para ello, se analiza un subconjunto de respuestas que dieron 192 docentes a un cuestionario en línea, desde los niveles de comprensión conceptual, bajo taxonomía SOLO, para caracterizar y ejemplificar el conocimiento de los temas estadísticos en juego por parte del profesorado participante.

2. Marco conceptual

El modelo de Conocimiento Especializado del Profesor de Matemática –MTSK, *Mathematics Teacher's Specialized Knowledge*– surge en coherencia y consistencia tanto teórica como práctica para examinar analíticamente las especificidades del conocimiento que manifiesta el profesorado que enseña matemática (Carrillo et al., 2013; Carrillo et al., 2018; Flores et al., 2013). Este modelo distingue el dominio de conocimiento matemático (MK, *Mathematical Knowledge*) y el dominio de conocimiento didáctico del contenido (PCK, *Pedagogical Content Knowledge*), además de las creencias sobre dichos dominios en entornos educativos diversos.

El modelo MTSK se basó en Shulman (1986, 1987) al considerar dos dominios de conocimiento profesional docente –i. e., el PCK y el SMK, *Subject Matter Knowledge*

(Conocimiento de la Materia)–. A diferencia del modelo MKT (Ball et al., 2008) –*Mathematical Knowledge for Teaching*– el modelo MTSK conceptualiza la noción de especialización, reconfigurando el conocimiento matemático y reinterpretando el conocimiento didáctico del contenido. En ese sentido, Scheiner et al. (2019) precisan que la *especialización* no puede explicarse de manera exhaustiva abordando lo que sabe el profesorado, sino también el cómo se produce el conocimiento en dicho profesorado, lo cual requiere no solo un trabajo investigativo de campo, en donde el conocimiento se manifiesta, sino que también un trabajo que indague en la formación que supone una producción de conocimientos especializados de los docentes.

El dominio MK que manifiesta un profesor de Matemática es considerado por el modelo MTSK como uno disciplinar propio de la matemática científica enmarcada en un contexto educativo, extendiendo el SMK. Mientras que el PCK es el que contempla los contenidos matemáticos en función de los procesos de enseñanza, aprendizaje y evaluación que se interrelacionan y determinan durante el quehacer docente al momento de educar matemáticamente a sus estudiantes.

En lo concerniente al modelo MTSK, la estadística no podrá ser concebida como una rama de la matemática, ya que la estadística es una ciencia independiente. En palabras de Cobb y Moore (1997), "al igual que la economía y la física, la estadística hace un uso intensivo y esencial de las matemáticas, sin embargo, tiene un territorio propio que explorar y conceptos centrales propios para guiar la exploración" (p. 814). Asimismo, Zieffler et al. (2018) afirman que la educación estadística ha desarrollado autonomía e independencia respecto de la educación matemática. En consecuencia, es oportuno precisar un modelo que extienda el modelo MTSK a la estadística, surgiendo así el Modelo de Conocimiento Especializado del Profesor de Estadística –STSK, *Statistics Teacher´s Specialised Knowledge*– propuesto por Vidal-Szabó y Estrella (2020).

El modelo STSK en ciernes reconoce, como hipótesis de trabajo, la existencia de un doble rol del profesorado que enseña matemática, pues también enseña estadística; por ello, se entenderá por "profesor de Estadística" al profesor de Matemática que enseña estadística, siendo habitual en los sistemas escolares (e. g., Australian Curriculum, Assessment and Reporting Authority, 2013; Common Core State Standards Initiative, 2017; New Zealand Ministry of Education, 2007). Asimismo, el término *conocimiento especializado* en el propuesto modelo STSK aplica a todo conocimiento que el profesorado requiera en su labor educativa profesional de enseñar Estadística, quedando excluidos los conocimientos referidos a la pedagogía general o los conocimientos de otros profesionales que emplean la estadística en sus campos laborales (no educativos).

El KoT-estadístico (*Knowledge of Topics*) es un tipo de conocimiento que refiere al qué y de qué forma el profesor de Estadística comprende los temas que enseña, en particular, involucra un conocimiento profundo de los conceptos como contenido estadístico, incluyendo intraconexiones en el tema (Vidal-Szabó y Estrella, 2020). A continuación, la Tabla 1 presenta una propuesta de categorías y descripciones del KoT-estadístico, en base al modelo MTSK y la literatura disponible en educación estadística (*e. g*., Pfannkuch, 2011; Vidal-Szabó et al., 2020; Wild y Pfannkuch, 1999; Wild et al., 2018).

Tabla 1 . Categorías y descripciones en relación al KoT-estadístico del profesorado. Nota. Elaboración propia.

Categoría	**Descripción**
A. Procedimientos: Características del resultado estadístico	Responde a cómo, cuándo y por qué hacer estadística. Por ejemplo, involucra saber procedimientos estadísticos convencionales y alternativos, también conocer características del resultado estadístico subyacente.
B. Definiciones, propiedades y fundamentos estadísticos	Para cualquier trabajo estadístico, el conocimiento de propiedades y principios estadísticos subyacentes es fundamental para entender el comportamiento de los datos.
C. Fenomenología y aplicaciones	Existe una gama de fenómenos o situaciones vinculada a ciertos temas estadísticos. Por ejemplo, el conocimiento del profesor de los diferentes contextos asociados con el concepto de moda y sus significados estadísticos. El conocimiento fenomenológico del profesor sobre el tema estadístico incluye los usos y aplicaciones de dicho tema.

2.1 Conceptos estadísticos elementales

En el KoT, la idea tras el término *Tema* es la especificidad del conocimiento que contempla la parte del currículo de Matemática que comprende contenidos estadísticos a tratar en la formación escolar. Específicamente, esta investigación aborda los conocimientos que expresan profesores en Chile que enseñan estadística en educación básica relacionados con los conceptos variable estadística, dato, información estadística y contexto de los datos; conceptos necesarios para el desarrollo de los modos de usar y pensar los datos con sentido. A continuación, se proponen descripciones para cada concepto.

Variable estadística. Es una característica medible (u observable) que adopta diferentes valores en un conjunto de individuos de una determinada población y que permite estudiar algún fenómeno de interés en un contexto específico. En un sentido amplio, se distinguen dos tipos de variables: cualitativas (nominal u ordinal) y cuantitativas (discreta o continua).

Dato. Es cualquier valor (numérico o categórico) que puede tomar una variable estadística al ser medida –por observación o mediante algún instrumento– en algún individuo de una determinada población. La palabra dato, del latín *datum*, significa algo dado, pues el dato corresponde a un individuo –unidad estadística relativo a una población–, el cual porta una información en un contexto determinado.

Información estadística. Es un conocimiento que se adquiere producto del análisis estadístico, extrayéndose durante y después de la recogida de datos. La información es producto de las interpretaciones plausibles sobre el comportamiento de los datos que provienen de una población o parte de ella.

Contexto de los datos. Tiene relación con la procedencia de los datos, la cual puede ser un entorno físico o una situación de tipo político, histórico, temporal, cultural o de otra índole. En estadística, los datos y su contexto están indisolublemente asociados al concebir los datos como números en un contexto específico.

Para efectos de este artículo, el foco está en el conocimiento estadístico y, en particular, en el subdominio del conocimiento de los temas (KoT-estadístico).

3. Metodología

3.1 Participantes y contexto

Esta investigación es de carácter cualitativo y pretende dar evidencias de la caracterización de un conocimiento especializado de temas estadísticos en profesores de Estadística en educación básica (tipificación otorgada al profesorado participante). La muestra no-probabilística por conveniencia y heterogénea abarcó 192 profesores que realizan clases en educación básica (1° a 6° año básico) en escuelas chilenas, contestando un cuestionario en línea el año 2020 y que, según orden cronológico, fueron rotulados desde P001 hasta P192. Estos docentes dieron su autorización sobre el uso de sus respuestas, cuidando una circulación reservada para fines de investigación y garantizando anonimato.

Respecto al perfil profesional, un 88% son profesores de educación básica y el resto de educación diferencial. La distribución de los docentes, según establecimiento educativo de tipo municipal, particular subvencionado y privado es de 133, 50 y 9 docentes, respectivamente. Asimismo, la zona de Chile donde más ha trabajado el profesorado participante corresponde a 41 docentes en la zona norte (desde la región de Arica y Parinacota hasta la región de Coquimbo), 126 docentes en la zona centro (desde la región de Valparaíso hasta la región del Maule) y 25 docentes en la zona sur (desde la región de Ñuble hasta la región de Magallanes y de la Antártica chilena).

3.2 Proceso de validación del instrumento de recogida de datos

El instrumento inicial fue sometido a un proceso de validación que permitió refinarlo por medio del método por juicio experto. Este método consiste en solicitar a personas un juicio hacia un instrumento y su opinión respecto a un aspecto específico, resultando útil en la valoración de aspectos netamente cualitativos; además, como técnica, su adecuada realización metodológica es un indicador de validez de contenido del instrumento de recogida de datos (Escobar y Cuervo, 2008).

El método por juicio experto permitió dar cuenta de la fiabilidad del instrumento, al establecer una opinión informada de personas expertas en el tema, quienes son reconocidas por otros como expertos y que pueden dar juicios y valoraciones. La selección de los 15 expertos consideró una semblanza breve de cada uno a partir de su trayectoria (experiencia y formación situadas en educación básica), infiriéndose su adecuada y pertinente participación en el rol de juez experto (ver anexo 1).

El cuestionario en línea, a través de Google Forms, permitió que los expertos realizaran sus juicios respecto al instrumento. El propósito de ello consistió en validar el contenido y la construcción de los ítems. En cada uno de los ítems, los jueces cumplieron con dos acciones:

> (i) Evaluar en una escala discreta desde 0 hasta 5 el grado de comprensión referido a qué tan comprensibles eran los elementos contenidos en el ítem para un docente; el grado de precisión referido a qué tan precisos eran los conceptos utilizados en el ítem, sin ambigüedad, y el grado de pertinencia referido a qué tan pertinente era el objetivo del ítem y las eventuales respuestas al mismo.
>
> (ii) Mejorar el ítem, según estimaran más comprensible, preciso y/o pertinente, en la sección “Observaciones o sugerencias”.

Una vez realizado el método de validación y examinado el grado de acuerdo de todos los jueces participantes, en base a las acciones (i) y (ii), el cuestionario se formuló vía Google Forms en su versión final. Esta herramienta en línea permitió aplicar 24 ítems con formatos de respuesta tales como varias opciones (42,3% ítems de selección única), casillas (11,5% ítems de selección múltiple), párrafo (30,8% ítems de respuesta escrita) y escala lineal (15,4% ítems de valoración). El cuestionario consta de secciones agrupadas: (a) consentimiento informado (sección 1); (b) información sobre su perfil profesional (sección 2 a la 10); (c) la estadística y su enseñanza (sección 11 a la 23); (d) interpretación gráfica (sección 24 a la 28); (e) representaciones de datos en educación básica (sección 29 a la 33), y (f) reflexiones finales y contacto (sección 34 a la 37).

Para efectos de este artículo, el análisis de las respuestas se enfoca en los ítems 5, 11, 6 y 14 cuyo formato es de respuesta escrita y referidos respectivamente a variable estadística, dato, información estadística y contexto de los datos. Estos ítems tuvieron un grado de acuerdo sobre 3 en la acción (i) y no hubo modificaciones mayores a estos en la acción (ii).

3.3 Procedimiento de análisis de las respuestas al cuestionario

La taxonomía SOLO (*Structure of Observed Learning Outcome*) de Biggs y Collis (1989) describe el incremento de la complejidad en el desempeño de tareas de aprendizaje, lo que brinda un enfoque para categorizar el rendimiento cognitivo teniendo en cuenta la estructura del resultado de aprendizaje observado, de modo que una respuesta viene a ser un resultado de aprendizaje que puede observarse, la cual es provocada por una pregunta. En ese sentido, la taxonomía SOLO postula cinco niveles ascendentes: (1) Preestructural (PE), el nivel más bajo, la respuesta que no ha captado la pregunta; (2) Uniestructural (UE), la respuesta dada al ítem capta solo una parte de la tarea; (3) Multiestructural (ME), la respuesta es solo una descripción cualitativa de la situación; (4) Relacional (R), la respuesta da cuenta que integra la descripción cualitativa con un aspecto cuantitativo; (5) Abstracto ampliado (A+) la respuesta integra lo cualitativo (Estrella et al., 2019). Nótese que el grado de complejidad en la respuesta depende tanto de la capacidad cognitiva del individuo como de la dificultad de la pregunta. El anexo 2 describe más detalladamente los niveles jerarquizados PE, UE, ME, R y A+.

Las respuestas a los ítems considerados fueron sometidas a un proceso de clasificación por medio de la taxonomía SOLO, puesto que permite examinar las respuestas por nivel para describirlas y caracterizarlas de acuerdo con la manifestación de algún conocimiento especializado referido a variable estadística, dato, información estadística y contexto de los datos. Dicho proceso consideró dos fases, en la primera, uno de los autores de la presente investigación hizo las clasificaciones de las 768 respuestas totales provenientes de los ítems 5, 6, 11 y 14 con la taxonomía SOLO. Luego, en la fase 2, conjuntamente los dos autores revisaron cada una de las respuestas clasificadas, según los niveles jerarquizados, ratificándose la mayoría y consensuando las discrepancias.

4. Análisis y resultados

A continuación, se presentan las clasificaciones de las respuestas entregadas por ítem por medio de la taxonomía SOLO y, posteriormente, se analizan algunas evidencias de acuerdo con la manifestación del KoT-estadístico en el profesorado participante.

4.1 Clasificación de las respuestas, según taxonomía SOLO

4.1.1 Análisis del ítem 5: Variable Estadística

Para la consigna: *En síntesis, dé un ejemplo para explicar lo que para usted es la "variable estadística"*, las 192 respuestas al ítem 5 se clasificaron según la descripción propuesta para este concepto pertinente al ítem y concordante a la taxonomía SOLO (Tabla 2).

Tabla 2 . Clasificación de las respuestas al ítem 5, según taxonomía SOLO. Nota. Elaboración propia.

Nivel y descripción	Frec.	Algunas respuestas
PE. Explicación incorrecta sin ejemplos, o bien, explicación con ejemplo(s) incorrecto(s) sobre variable estadística.	45 (23,44%)	P009. Es un dato que depende de otros. P062. Información sobre una característica de la población a estudiar. P071. Medir matemáticas distintas situaciones. P100. Es una rama de la matemática. P122. En el día a día en cada momento de nuestras vidas se enseña la variable estadística ya que se toma en cuenta cada una de las variables para profundizar las etapas de nuestras vidas. P172. Recoger información y transmitirla como datos, gráficos y comparar dicha información.
UE. Explicación parcialmente correcta sin ejemplo(s), o bien, solo ejemplo(s) sin explicación sobre variable estadística.	111 (57,81%)	P010. Que adoptan valores numéricos, edad, peso. P119. ¿Cuál es tu etnia? Mapuche, Rapa-Nui, Aymara. P143. Color de las bebidas gaseosas: rojo, amarillo, negro, naranja. P170. Número de hermanos (cuantitativa), estado civil (cualitativa). P189. Personas contagiadas con COVID. P191. Es una variable de la cual se quiere entender su tendencia. Ej.: las evaluaciones de los alumnos.
ME. Explicación mayormente correcta sin ejemplo(s), o bien, explicación con ejemplo(s) parcialmente correcta sobre variable estadística.	33 (17,19%)	P005. Una variable estadística es cada una de las características o cualidades que poseen individuos en una población, como por ejemplo color de pelo en el 4° básico del colegio de Talca. P034. Lo que se estudia de un fenómeno y que tiene distintas opciones. Por ejemplo, color de ojos. P050. Característica cualitativa o cuantitativa que puede ser medible u observable.
R. Explicación correcta sobre la variable estadística con o sin ejemplo(s).	3 (1,60%)	P093. Es una característica cambiante que puede medirse... Por ejemplo, la estatura, el color de pelo. P115. [...] que puede cambiar, se puede observar y se puede medir. Las variables del cambio de ánimo en tiempos de pandemia son: miedo, encierro, falta de recursos, pérdida de fuente laboral. P187. [...] puede ser numérica (cuantitativa) o no numérica (cualitativa), que, al obtenerse de la muestra, permite saber información sobre ella. Por ejemplo, si yo quiero saber cuál es el color favorito de una persona para hacer el logo de una marca, la variable sería el color favorito (que sería cualitativa) y le preguntaría a una cierta cantidad de personas su color favorito y el que tenga mayor cantidad de votos iría dentro del logo.
A+. Explicación correcta con o sin ejemplo(s) y que incluye conceptos que amplían el significado de variable estadística.	0 (0%)	.

La mayoría de las respuestas al ítem 5 pertenecen al nivel UE, cuya principal característica es tener como respuesta sobre variable estadística una explicación parcialmente correcta sin ejemplo(s), como lo hacen P010 y P191, entre otros; o bien, solo un(os) ejemplo(s) sin explicación, como lo hacen P119 y P170, entre otros. Mientras que, en la segunda mayoría, un poco menos del cuarto de las respuestas pertenecen al nivel PE (23,44%; 45 docentes), siendo estas incorrectas respecto a la consigna del ítem 5.

Además, un poco menos del 2% de las respuestas están en el nivel R (solo 3 docentes). Dichas respuestas van más allá del tipo multiestructural, pues la explicación sobre la variable estadística está operacionalizada con o sin ejemplo(s) y es correcta. P093 ilustra lo anterior, pues dice que "es una característica cambiante que puede medirse [...]", vinculando la variable estadística al cambio que puede ser medido. En cambio, P187 da un ejemplo específico para explicar variable estadística como sigue: "[...] si yo quiero saber cuál es el color favorito de una persona para hacer el logo de una marca, la variable sería el color favorito (que sería cualitativa) y le preguntaría a una cierta cantidad de personas su color favorito y el que tenga mayor cantidad de votos iría dentro del logo", situando la variable estadística para una toma de decisión, a través del sondeo que señala. Nótese que no hay respuestas categorizadas como A+.

4.1.2 Análisis del ítem 11: Dato

Para la consigna: *En síntesis, dé un ejemplo para explicar lo que para usted es un dato en estadística*, las 192 respuestas al ítem 11 se clasificaron según la descripción propuesta para este concepto pertinente al ítem y concordante a la taxonomía SOLO (Tabla 3).

Tabla 3.Clasificación de las respuestas al ítem 11, según taxonomía SOLO

Nota. Elaboración propia.

Nivel y descripción	**Frec.**	**Algunas respuestas**
PE. Explicación incorrecta sin ejemplos, o bien, explicación con ejemplo(s) incorrecto(s) sobre dato estadístico.	119 (61,98%)	P023. Es recoger información sobre una realidad del mundo y sus intereses. P049. Un número. P056. Por ejemplo, poder realizar la variación de la matrícula en la escuela... O separar por género la matrícula. P070. Cantidad de estudiantes ausentes los días viernes en cuarto básico. P134. Color de pelo, estatura, cantidad de basura que aporta cada región, tipos de basura que existen, etc. P172. Recoger información en un contexto.
UE. Explicación parcialmente correcta sin ejemplo(s), o bien, solo ejemplo(s) sin explicación sobre dato estadístico.	24 (12,5%)	P040. Un dato en estadística es un valor. P046. La edad de uno de mis alumnos. P068. Edades: Juan 8 años, Amanda 7 años, Alicia 7 años, Moisés 9 años. (Cada edad de los niños es un dato estadístico). P072. Al tirar un dado 8 veces obtengo 8 dados. P166. Pelo liso pelo liso pelo liso pelo ondulado pelo liso pelo ondulado.
ME. Explicación mayormente correcta sin ejemplo(s), o bien, explicación con ejemplo(s) parcialmente correcta sobre dato estadístico.	46 (23,96%)	P007. Son los valores obtenidos al realizar un estudio. P009. Una cantidad en un contexto. P014. Son los valores, cuántos autos rojos y azules hay en el estacionamiento. P029. Es una cifra o un elemento de la medición obtenida de las estadísticas. P075. Es cada uno de los valores que se ha obtenido al realizar un estudio estadístico. Por ejemplo, el 1 es mujer y el 2 es varón. P085. Es el valor obtenido al hacer un estudio estadístico, por ejemplo, si lanzamos un dado 3 veces los datos serían que una vez salió 5, otra 3 y otra 1. P148. Un dato es una característica cuantificable y medible.
R. Explicación correcta sobre el dato estadístico con o sin ejemplo(s).	1 (0,52%)	P045. Por ejemplo, si juego 8 veces Cachipún con una persona, puedo obtener 8 datos (papel, tijera, papel, papel, piedra, papel, piedra y papel). Con esto puedo ver que el tiro más utilizado por mi contrincante es el papel.
A+. Explicación correcta con o sin ejemplo(s) y que incluye conceptos que amplían el significado de dato estadístico.	2 (1,04%)	P034. Un resultado de una variable, por ejemplo, variable notas en la prueba, dato 6.5. P115. Son cada una de las posibles respuestas que se puede obtener en un estudio, ejemplo: si pateo 3 penales obtengo datos: gol-gol-fuera.

La mayoría de las respuestas al ítem 11 están en el nivel PE (61,98%; 119 docentes), esto es, un poco más de los tres quintos de las respuestas son incorrectas, pues están vinculadas a las ideas de número sin contexto, como en P049; están relacionadas a la información indistinguiéndola del dato, como en P023 y P172; toman el concepto de variable estadística pero no la precisan de acuerdo con el dato, como en P056 y P070. Mientras que, en la segunda mayoría, un poco menos de una cuarta parte de las respuestas pertenecen al nivel ME (23,96%; 46 docentes), siendo mayormente correctas. Entre ellas, se rescata principalmente la idea sobre el valor que toma la variable, como en P007, P014 y P085, entre otras respuestas.

Muy pocas respuestas están en las categorías R o A+. Por ejemplo, en la categoría R, P045 separa el concepto de variable del concepto de dato, pues indica que "[...]

si juego 8 veces Cachipún con una persona, puedo obtener 8 datos (papel, tijera, papel, papel, piedra, papel, piedra y papel)". Mientras que en las respuestas tipo A+, P115 en el contexto del fútbol otorga una respuesta situada en un estudio que contiene tanto la incertidumbre ("posibles respuestas") como la variable estadística con los valores que toma ("gol-gol-fuera"). En cambio, P034 en el contexto educativo de las notas (calificaciones) da una respuesta precisa con ejemplo correcto, ya que señala: "Un resultado de una variable" y da de ejemplo para una variable como las notas en una prueba el dato "6.5", que corresponde a la calificación de un estudiante en dicho contexto.

4.1.3 Análisis del ítem 6: Información estadística

Para la consigna: *En síntesis, dé un ejemplo para explicar lo que para usted es la "información estadística",* las 192 respuestas al ítem 6 se clasificaron según la descripción propuesta para este concepto pertinente al ítem y concordante a la taxonomía SOLO (Tabla 4).

Tabla 4 . Clasificación de las respuestas al ítem 6, según taxonomía SOLO. Nota. Elaboración propia.

Nivel y descripción	**Frec.**	**Algunas respuestas**
PE. Explicación incorrecta sin ejemplos, o bien, explicación con ejemplo(s) incorrecto(s) sobre información estadística.	154 (80,21%)	P004. Color de las bebidas gaseosas: amarillo, negro, naranja, contenidos de las bebidas; 0,5 litros; 1,5 litros; 2,5 litros. P006. Son los datos obtenidos de un estudio realizado, los que se ocupan para tomar decisiones. P017. Grupo de datos de una población. P038. Son todos los datos que se buscan recoger para luego ser analizados. P139. Rama de las matemáticas que estudia las variables. P159. Es una especialidad de la matemática para reflejar cuantitativamente un fenómeno. P165. Pienso que es la realidad medible y entendible en base a una mirada lógica. P182. Por ejemplo, si lanzamos una moneda al aire 5 veces obtenemos 5 datos. Eso es una información estadística.
UE. Explicación parcialmente correcta sin ejemplo(s), o bien, solo ejemplo(s) sin explicación sobre información estadística.	21 (10,94%)	P057. En una encuesta de 100 personas 34% son casados, 25% separados, 27% solteros y 14% viudos. P068. En 4to básico, 20 de 35 estudiantes tienen un perro de mascota. P110. El 80% de los estudiantes de primero básico consume dos tipos de frutas (peras y manzanas) durante el 2019. P166. 20 tienen pelo liso y 10 tienen ondulado.
ME. Explicación mayormente correcta sin ejemplo(s), o bien, explicación con ejemplo(s) parcialmente correcta sobre información estadística.	11 (5,73%)	P009. Corresponde a la interpretación que se les da a los datos. P080. Son los resultados que entrega una investigación.
R. Explicación correcta sobre información estadística con o sin ejemplo(s).	3 (1,56%)	P094. Corresponde a la información entregada a través de datos cuantitativos ordenados, lo que permite el análisis, la reflexión y la toma de decisiones. Esta información puede ser manipulada. P109. Está constituida por un grupo de datos ya supervisados y ordenados, que sirven para construir un mensaje basado en un cierto fenómeno o ente. La información permite resolver problemas y tomar decisiones, ya que su aprovechamiento racional es la base del conocimiento. P125. La información estadística es la que resume, organiza e intenta simplificar un conjunto de datos de estilo numeroso o muy complicado.

A+. Explicación correcta con o sin ejemplo(s) y que incluye conceptos que amplían el significado de información estadística.	3 (1,56%)	P092. Es el análisis que se le puede dar al resultado del estudio de una variable estadística. P131. Corresponde al análisis y conclusiones extraídos de los datos relevantes de un estudio estadístico. P192. Es el mensaje que entrega el trabajo estadístico realizado, tomando en cuenta las variables aplicadas en un contexto determinado.

En el ítem 6, la gran mayoría de las respuestas están en el nivel PE (80,21%; 154 docentes), cuya principal característica es definir el concepto información estadística de manera inadecuada y/o dar un ejemplo errado. Por ejemplo, P004 ("color de las bebidas gaseosas: amarillo, negro, naranja, contenidos de las bebidas; 0,5 litros; 1,5 litros; 2,5 litros") confunde información estadística con el concepto de variable estadística y los valores que puede tomar. Asimismo, P159 indica que la información estadística "es una especialidad de la matemática para reflejar cuantitativamente un fenómeno", lo cual evidencia la indistinción entre matemática y estadística.

Un poco menos del 20% de las respuestas (38 docentes) están en los niveles sobre PE. Las respuestas que pertenecen al nivel UE son pocas (10,94%; 21 docentes) y dan cuenta de un ejemplo específico para información estadística. Por ejemplo, P068 señala: "en 4to básico, 20 de 35 estudiantes tienen un perro de mascota", P110 es aún más específico de acuerdo con el contexto, "el 80% de los estudiantes de primero básico consume dos tipos de frutas (peras y manzanas) durante el 2019". En cambio, las respuestas que están en el nivel ME esbozan definiciones más generales, por ejemplo, P009 responde que la información estadística: "corresponde a la interpretación que se les da a los datos". Hay tres respuestas por nivel en R y en A+, que dan cuenta de una comprensión más acabada sobre información estadística. Por ejemplo, en el nivel R, P094 indica que "corresponde a la información entregada a través de datos cuantitativos ordenados, lo que permite el análisis, la reflexión y la toma de decisiones. Esta información puede ser manipulada"; y en el nivel A+, P192 indica que la información estadística "es el mensaje que entrega el trabajo estadístico realizado, tomando en cuenta las variables aplicadas en un contexto determinado".

4.1.4 Análisis del ítem 14: Contexto de los datos

Para la consigna: *En síntesis, dé un ejemplo para explicar lo que para usted es el contexto de los datos*, las 192 respuestas al ítem 14 se clasificaron según la descripción propuesta para este concepto pertinente al ítem y concordante a la taxonomía SOLO (Tabla 5).

Tabla 5 . Clasificación de las respuestas al ítem 6, según taxonomía SOLO. Nota. Elaboración propia.

Nivel y descripción	**Frec.**	**Algunas respuestas**
PE. Explicación incorrecta sin ejemplos, o bien, explicación con ejemplo(s) incorrecto(s) sobre el contexto de los datos.	94 (48,96%)	P001. Es el mínimo de información recolectada. P006. Es la información que se obtiene y que nos permite cuantificar un dato que se quiere averiguar, por ejemplo, terneros nacidos vivos en una parcela en el mes de junio. P010. Datos, comparados, analizados e interpretados.
UE. Explicación parcialmente correcta sin ejemplo(s), o bien, solo ejemplo(s) sin explicación sobre el contexto de los datos.	53 (27,60%)	P031. Por ejemplo, la pandemia que estamos viviendo hoy en día, los datos estadísticos son verídicos. P046. Trabajar con los números de calzado dados por mis alumnos para trabajar cuánto calzan en promedio. P060. Obtener datos del entorno del barrio y familia de los estudiantes. Datos verdaderos que los niños puedan obtener y verificar. P069. A nivel regional por ejemplo cuál es el ingreso per cápita por persona en las diferentes regiones de Chile. P081. Cantidad de personas con oficios y profesionales en las diferentes clases sociales. P088. A quién va dirigida la entrevista… Ej.: abuelos. P123. Lo que se desea saber de los estudiantes, ej.: qué alimento es el que más gusta. P131. La diferencia entre un curso a otro, los datos pueden ser distintos. P189. A quiénes se aplicó el examen, en qué momento, hubo contactos…

ME. Explicación mayormente correcta sin ejemplo(s), o bien, explicación con ejemplo(s) parcialmente correcta sobre el contexto de los datos.	36 (18,75%)	P002. De donde se obtienen. P020. Es el lugar donde se originan o de donde provienen los datos. P035. Es muy diferente consultar sobre alimentación en la sociedad chilena actual, sin tomar en cuenta el contexto donde hago dicha encuesta. P040. Escenario de donde proceden. P092. El lugar en donde se recogen los datos, por ejemplo: el curso en el que se aplica una encuesta. P095. La realidad de donde se obtienen. P113. Es el medio del cual se recogen. P155. El contexto es fundamental para que el dato sea útil y valioso. Por ejemplo: Cantidad de cajas de mercadería entregadas por comunas de la Región Metropolitana en los meses de emergencia sanitaria por COVID-19.
R. Explicación correcta sobre el contexto de los datos con o sin ejemplo(s).	6 (3,13%)	P026. Tiene que ver con el ambiente en que se toman los datos, por ejemplo, los datos que se obtienen en el logro de los objetivos en una escuela particular versus una municipal o particular subvencionado. P102. La realidad que rodea un dato numérico o no numérico P115. Tiene que ver con el momento y lugar donde se recogen los datos. Ej., si yo recojo datos sobre agrado de las matemáticas en un alumno después de sacarse una nota 7,0. sería diferente su respuesta si le pregunto después de haberse sacado una nota 2,0. P170. Si realizo un estudio sobre horas de estudio que dedica un estudiante para lo anterior debo considerar el contexto en el que viven o se desenvuelven mis sujetos de estudio, ya que distintos datos obtendré si este estudio lo realizo en un sector acomodado a un sector de mucha vulnerabilidad social (por lo tanto, es importante el contexto en el cual obtuve mis datos y si son de utilidad para el objetivo de mi estudio) P174. En qué periodo de tiempo y cómo se obtuvieron los datos. La descripción de los encuestados. Por ejemplo: Se realizó una encuesta telefónica a personas entre 15 y 25 de la Región Metropolitana.
A+. Explicación correcta con o sin ejemplo(s) y que incluye conceptos que amplían el significado del contexto de los datos.	3 (1,56%)	P048. Entorno que condiciona los datos. P051. Es lo que rodea al dato. Por ejemplo, el periodo de obtención de los datos a estudiar o a qué persona corresponden estos datos.

En el ítem 14, un poco menos de la mitad de las respuestas están en el nivel PE (48,96%; 94 docentes). Por ejemplo, P006 indica que el contexto de los datos “es la información que se obtiene [...]”, lo cual es impreciso, pues no corresponde al contexto de los datos; en tanto, P010 indica que son “datos, comparados, analizados e interpretados”, respuesta que no diferencia el contexto de los datos con el proceso estadístico propio en una investigación empírica. En el nivel UE, las respuestas corresponden mayormente a ejemplos específicos, como “la diferencia entre un curso a otro, los datos pueden ser distintos” (P131). Y en el nivel ME existen explicaciones que contemplan una definición, por ejemplo: “Es el lugar donde se originan o de donde provienen los datos” (P020); “El lugar en donde se recogen los datos, por ejemplo: el curso en el que se aplica una encuesta” (P092). Dichas respuestas van más allá de un ejemplo específico, tienen incorporada la idea del contexto en relación con la procedencia de los datos, pudiendo ser un entorno físico o una situación de tipo político, histórico, temporal, cultural o de cualquier otra índole.

Pocas respuestas están en el nivel R, o bien A+, alcanzando un 4,7% de las respuestas aproximadamente. En el caso de las respuestas en el nivel R, contemplan una explicación que contienen una definición más precisa y, a veces, la operacionalizan con un ejemplo correctamente como lo ilustra P026, quien indica: “Tiene que ver con el ambiente en que se toman los datos, por ejemplo, los datos que se obtienen en el logro de los objetivos en una escuela particular versus una municipal o particular subvencionado”; en cambio, P174 señala: “En qué periodo de tiempo y cómo se obtuvieron los datos. La descripción de los encuestados. Por ejemplo: Se realizó una encuesta telefónica a personas entre 15 y 25 de la Región Metropolitana”. En el caso de las respuestas en el nivel A+ existe una ampliación conceptual sobre el contexto de los datos, por ejemplo, P051 señala que el contexto de los datos “es lo que rodea al dato. Por ejemplo, el periodo de obtención de los datos a estudiar o a qué persona corresponden estos datos”; P048 enuncia que el contexto es el “entorno que condiciona los datos”. Este último da cuenta de cómo el contexto provee en los datos un criterio por el cual validarlos, en tanto están condicionados los datos al contexto.

4.1.5 Análisis de los conceptos estadísticos clasificados, según taxonomía SOLO

El gráfico de la Figura 1 muestra el comportamiento de las respuestas para cada uno de los conceptos estadísticos, considerando la clasificación dada por los niveles de la taxonomía SOLO.

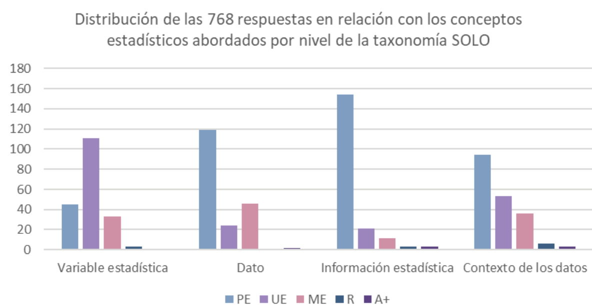


Figura 1 . Gráfico de la distribución de respuestas referidas a los conceptos estadísticos abordados, según taxonomía SOLO
Nota. Elaboración propia.

Notar que son escasas las respuestas que están categorizadas en los niveles ME, R o A+, en comparación a los niveles PE y UE que son los que predominan. Además, el concepto de información estadística es el más descendido respecto a su comprensión, posiblemente porque es confundido por variable estadística o por dato u otro concepto.

4.2 Evidencias de la manifestación del KoT-estadístico

A continuación, se presentan algunas evidencias de la manifestación del KoT-estadístico en profesores de Estadística que enseñan en educación básica, tomando en cuenta la clasificación de las respuestas a los ítems 5, 6, 11 y 14, según los niveles ME, R y A+ de la taxonomía SOLO.

(A) Procedimientos: característica del resultado estadístico. P187, al explicar el concepto de variable estadística, señala: "[...] si yo quiero saber cuál es el color favorito de una persona para hacer el logo de una marca, la variable sería el color favorito (que sería cualitativa) y le preguntaría a una cierta cantidad de personas su color favorito y el que tenga mayor cantidad de votos iría dentro del logo", ello indica que identifica el tipo de variable y establece una manera de proceder para determinar el color favorito de una persona para la elaboración del logo de una marca. En cambio, P170, para el concepto del contexto de los datos enuncia que: "[...] debo considerar el contexto en el que viven o se desenvuelven mis sujetos de estudio, ya que distintos datos obtendré si este estudio lo realizo en un sector acomodado a un sector de mucha vulnerabilidad social [...]", con lo que sitúa el contexto de los datos como un factor relevante en la recolección de datos, siendo además una fuente de variabilidad. Estos docentes dan evidencias de un KoT-estadístico porque están involucrando en sus explicaciones ciertos procedimientos convencionales propios de un trabajo estadístico, los cuales pueden estar relacionados con sus conocimientos sobre la práctica de los estadísticos (o usuarios de la estadística) en una investigación empírica.

(B) Definiciones, propiedades y fundamentos estadísticos. P093 señala una definición para variable estadística como "[...] una característica cambiante que puede medirse [...]"; también P005 enuncia que "una variable estadística es cada una de las características o cualidades que poseen individuos en una población [...]", estos docentes articulan en su explicación una definición adecuada. Para el caso del dato, P009 indica que es "una cantidad en un contexto", P034 señala que el dato es "un resultado de una variable, por ejemplo, variable notas en la prueba, dato 6.5"; estos docentes definen dato considerando su procedencia y su relación con alguna variable. Por su parte, para el caso de información estadística, P192 explica que es el "[...] mensaje que entrega el trabajo estadístico realizado, tomando en cuenta las variables aplicadas en un contexto determinado", distinguiendo información respecto de datos y de variable. Mientras que, para el caso del contexto de los datos, P102 lo define como "la realidad que rodea un dato numérico o no numérico", lo cual está próximo al conocimiento contextual relacionado a la variable estadística expresada en los datos. Estos docentes muestran evidencias de KoT-estadístico, ya que explican los conceptos por medio de definiciones adecuadas que construyen coherentemente en base a las ideas estadísticas fundamentales.

(*C) Fenomenología y aplicaciones.* P045, al explicar dato, ejemplifica: "[...] si juego 8 veces Cachipún con una persona, puedo obtener 8 datos (papel, tijera, papel, papel, piedra, papel, piedra y papel). Con esto puedo ver que el tiro más utilizado por mi contrincante es el papel". En este ejemplo, P045 reconoce los datos en una situación lúdica e incluye el concepto de moda, que puede ser útil como una estrategia en el juego. También, P115 para el caso del contexto de los datos da un ejemplo: "[...] si yo recojo datos sobre el agrado de las matemáticas en un alumno después de sacarse una nota 7,0 sería diferente a su respuesta si le pregunto después de haberse sacado una nota 2,0", lo que da cuenta de cómo el contexto incide en la variabilidad de los datos. Estos docentes dan evidencias de un KoT-estadístico, pues están vinculando ciertos temas a fenómenos o situaciones susceptibles de ser abordados con estadística.

5. Conclusiones

El propósito de la investigación fue indagar sobre los conocimientos y niveles de comprensión que expresaron profesores que enseñan estadística en educación básica sobre los conceptos de variable estadística, dato, información estadística y contexto de los datos. La idea del término *Tema* remite a los contenidos estadísticos que están considerados en el currículo, en el que dichos conceptos son temas fundamentales en la estadística escolar, por tanto, es relevante caracterizar estos temas en virtud de los conocimientos de los profesores de Estadística.

Los ejemplos revisados y previamente clasificados según la taxonomía SOLO –esto es, explicaciones que dan cuenta de los temas en cuestión y que fueron abordados por los docentes en los cuatro ítems del cuestionario en línea– respaldan la manifestación del KoT-estadístico en profesores que enseñan estadística en educación básica y es un conocimiento especializado, en tanto constituye un conjunto de subconocimientos que puede ser llevado a su actuar profesional docente en la educación estadística. De acuerdo con la clasificación realizada y el KoT-estadístico, fueron ejemplificadas las tres categorías en el modelo STSK, pudiendo servir como referencia para la formación de profesores de Estadística, pues permiten precisar y profundizar los conocimientos estadísticos especializados en el contexto educativo.

Los conceptos sobre variable estadística, dato, información estadística y contexto de los datos son necesarios para el desarrollo del sentido del dato y del razonamiento estadístico. Por tanto, dichos conceptos son fundamentales en la formación del profesor que enseña estadística en la escuela. Dada la importancia de generar oportunidades de aprendizaje desde los primeros años escolares, es necesario que el profesorado pueda comprender el sentido del dato que permite resolver problemas estadísticos a través de datos reales en contextos motivadores o de interés, comprender cómo se generan esos datos, elaborar juicios e inferencias a partir de la variabilidad de los datos y mostrar una postura crítica frente a la ingente información estadística.

La clasificación presentada de las respuestas a los cuatro ítems de los profesores que enseñan estadística en educación básica, da cuenta que todavía es un desafío la comprensión de los conceptos estadísticos abordados (ver Figura 1), pese a que se asume una comprensión de los términos estadísticos de uso común. Al respecto, en la línea de Watson y Fitzallen (2021), es posible que el profesorado no esté distinguiendo suficientemente a los datos como una expresión de la variable en el contexto de las investigaciones estadísticas, ni tenga en cuenta su carácter numérico o categórico de naturaleza medible u observable.

En Chile, el eje temático *Datos y Probabilidades* solo abarca aproximadamente un 12,6% promedio del tiempo escolar dedicado a la asignatura Matemática; sumado a lo anterior, unánimemente los objetivos de aprendizajes asociados a los contenidos estadísticos se encuentran en la última unidad propuesta en los programas de estudio de 1° a 6° año básico. Asimismo, en período de pandemia es preocupante que la priorización curricular haya afectado negativamente al currículo propio de la estadística en la asignatura Matemática, disminuyéndolo aún más. Por lo tanto, es necesario potenciar las experiencias de enseñar estadística en educación básica, lo cual podría permitir más oportunidades de desarrollo profesional en relación con los conocimientos especializados de los profesores de Estadística.

Se proyecta profundizar en los dominios de conocimiento del modelo STSK, ampliando tanto el conocimiento estadístico como el conocimiento didáctico del contenido de los profesores de Estadística, para contribuir al estudio del sentido del dato como precursor del aprendizaje estadístico y necesario en la desafiante formación del profesor de Estadística en el siglo XXI.

Agradecimientos

Esta investigación ha sido financiada parcialmente por ANID a través del Proyecto FONDECYT Nº 1200346 y CONICYT-PCHA / Doctorado Nacional: 2016-21161569. También ha sido patrocinada por el Grupo de Investigación en Estadística Temprana (GIET, https://estadisticatemprana.cl/) y el proyecto *Sumo Primero en Terreno* impulsado por la División de Educación General del MINEDUC en convenio con la Pontificia Universidad Católica de Valparaíso.

Referencias

Australian Curriculum, Assessment and Reporting Authority. (2013). *The Australian curriculum*. Author. https://www.australiancurriculum.edu.au/about-the-australian-curriculum/

Ball, D. L., Thames, M. H., y Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407. https://doi.org/10.1177/0022487108324554

Biggs, J., y Collis, K. F. (1989). Hacia un modelo de desarrollo y evaluación curricular basado en la escuela utilizando la taxonomía SOLO. *Australian Journal of Education, 33*, 151-163. https://doi.org/10.1177/168781408903300205

Burrill, G., y Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burril y C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education* (pp. 57-69). Springer. https://doi.org/10.1007/978-94-007-1131-0_10

Carrillo, J., Climent, N., Contreras, L. C., y Muñoz-Catalán, M. C. (2013). Determining specialised knowledge for mathematics teaching. En B. Ubuz, Ç. Haser y M. A. Mariotti (Eds.), *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education* (pp. 2985-2994). Middle East Technical University. http://www.mathematik.uni-dortmund.de/~erme/doc/CERME8/CERME8_2013_Proceedings.pdf#page2985

Carrillo, J., Climent, N., Montes, M., Contreras, L., Flores-Medrano, E., Escudero-Ávila, D., Vasco, D., Rojas, N., Flores, P., Aguilar-González, A., Ribeiro, M., y Muñoz-Catalán, M. C. (2018). The mathematics teacher's specialised knowledge (MTSK) model. *Research in Mathematics Education*, 1-18. https://doi.org/10.1080/14794802.2018.1479981

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

Common Core State Standards Initiative. (2017). *Mathematics standards*. http://www.corestandards.org/Math/

Del Pino, G., y Estrella, S. (2012). Educación Estadística: relaciones con la matemática. *Revista de Investigación Educacional Latinoamericana, Pensamiento Educativo, 49*(1), 53-64. https://doi.org/10.7764/PEL.49.1.2012.5

Escobar, J., y Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición, 6*, 27-36. https://www.researchgate.net/publication/302438451_Validez_de_contenido_y_juicio_de_expertos_Una_aproximacion_a_su_utilizacion

Estrella, S. (2017). Enseñar estadística para alfabetizar estadísticamente y desarrollar el razonamiento estadístico. En A. Salcedo (Ed.), *Alternativas Pedagógicas para la Educación Matemática del Siglo XXI* (pp. 173-194). Centro de Investigaciones Educativas, Escuela de Educación. Universidad Central de Venezuela. http://saber.ucv.ve/bitstream/123456789/15712/1/Alternativas%20Pedagogicas%20para%20la%20Educaci%C3%B3n%20Matematica%20S%20XXI.pdf

Estrella, S. (2018). Data representations in Early Statistics: data sense, meta-representational competence and transnumeration. En A. Leavy, M. Meletiou y E. Paparistodemou (Eds.), *Statistics in Early Childhood and Primary Education – Supporting early statistical and probabilistic thinking* (pp. 239-256). Springer. https://doi.org/10.1007/978-981-13-1044-7_14

Estrella, S., Alvarado, H., Olfos, R., y Retamal, L. (2019). Desarrollo de la alfabetización probabilística: textos argumentativos de estudiantes (según niveles de razonamiento de la taxonomía SOLO). *Revista Paradigma, 40*(1), 280-304.

Estrella, S., Vergara, A., y González, O. (en prensa). El desarrollo del sentido del dato: haciendo inferencias desde la variabilidad de los tsunamis en primaria. *Statistics Education Research Journal.*

Estrella, S., Zakaryan, D., Olfos, R., y Espinoza, G. (2020). How teachers learn to maintain the cognitive demand of tasks through Lesson Study. *Journal of Mathematics Teacher Education, 23*, 293-310. https://doi.org/10.1007/s10857-018-09423-y

Flores, E., Escudero, D. I., y Carrillo, J. (2013). A theoretical review of specialized content knowledge. En B. Ubuz, C. Haser y M. A. Mariotti (Eds.), *Proceedings of CERME8* (pp. 3055-3064). Middle East Technical University, Ankara. https://www.researchgate.net/publication/260266162_A_theoretical_review_of_Specialized_Content_Knowledge

Garfield, J., y Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* Springer Science y Business Media. https://doi.org/10.1007/978-1-4020-8383-9

Konold, C., Higgins, T., Russell, S., y Khalil, K. (2015). Datos vistos a través de diferentes lentes. *Estudios Educativos en Matemáticas, 88*, 305-325. https://doi.org/10.1007/s10649-013-9529-8

Ministerio de Educación de Chile. (2012). Matemática. En *Bases Curriculares para la Educación Básica* (pp. 85-135). http://www.curriculumenlineamineduc.cl/605/articles-21321_programa.pdf

New Zealand Ministry of Education. (2007). *The New Zealand curriculum.* Learning Media. https://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning, 13*(1 y 2), 27-46. https://doi.org/10.1080/10986065.2011.538302

Scheiner, T., Montes, M. A., Godino, J. D., Carrillo, J., y Pino-Fan, L. R. (2019). What makes mathematics teacher knowledge specialized? Offering alternative views. *International Journal of Science and Mathematics Education, 17*(1), 153-172. https://doi.org/10.1007/s10763-017-9859-6

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14. https://doi.org/10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review, 57*(1), 1-23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Vidal-Szabó, P., y Estrella, S. (2020). Extensión del modelo MTSK al dominio estadístico. En Y. Morales-López y Á. Ruíz (Eds.), *Educación Matemática en las Américas 2019* (pp. 1036-1042). Comité Interamericano de Educación Matemática. https://conferencia.ciaem-redumate.org/index.php/xvciaem/xv/paper/viewFile/692/327

Vidal-Szabó, P., Kuzniak, A., Estrella, S., y Montoya, E. (2020). Análisis cualitativo de un aprendizaje estadístico temprano con la mirada de los espacios de trabajo matemático orientado por el ciclo investigativo. *Revista Educación Matemática, 32*(2), 217-246. https://doi.org/10.24844/EM3202.09

Watson, J., y Fitzallen, N. (2021). What sense do children make of "data" by Year 3? En Y. H. Leong, B. Kaur, B. H. Choy, J. B. W. Yeo y S. L. Chi (Eds.), *Proceedings of the 43rd Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 409-416). The Mathematics Education Research Group of Australasia.

Wild, C., y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Wild, C., Utts, J., y Horton, N. (2018). What Is Statistics? En D. Ben-Zvi, K. Makar y J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 5-36). Switzerland: Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_2

Zieffler, A., Garfield, J., y Fry, E. (2018). What is Statistics Education? En D. Ben-Zvi, K. Makar y J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 37-70). Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_2

Anexo 1
Perfil de los 15 jueces expertos seleccionados.

Juez	Perfil
01	Prof. Gral. Básica con mención en Matemática y Mg. en Didáctica de la Matemática, mención ed. bás.
02	Profesora de Educación Básica y Magíster en Didáctica de la Matemática (DM).
03	Estudiante de 4° año de Ped. en Mat., mención en Estadística Educacional con destacada trayectoria.
04	Prof. de Educación Básica, Mg. en DM y candidata a Doctora en DM.
05	Profesor de Estadística y Candidato a Doctor en Didáctica de la Matemática.
06	Doctora en Didáctica de la Matemática.
07	Magíster en Evaluación y Currículo Educacional.
08	Profesora de Educación Básica y Magíster en Didáctica de la Matemática.
09	Profesor de Matemática y Computación. Experiencia en proyecto aRPa.
10	Doctor en Didáctica de la Matemática.
11	Profesora de Matemática y Magíster en Estadística.
12	Profesora de Matemática y Candidata a Magíster en Evaluación Educacional.
13	Magíster en Didáctica de la Matemática.
14	Doctora experimentada en Didáctica de la Matemática.
15	Magíster en Docencia de las Matemáticas y Lic. en Educación Básica con énfasis en Matemáticas.

Nota. Elaboración propia.

Anexo 2
Niveles jerarquizados de la taxonomía SOLO.

Nivel	Descripción
(PE) Preestructural,	Es el nivel más bajo, la respuesta da cuenta que el individuo no ha captado la pregunta. En otras palabras, se utiliza un dato o proceso incorrecto de forma simplista que puede llevar a una conclusión irrelevante. El individuo puede incluso no comprometerse con el problema, por lo que no hay ningún tipo de cierre en la respuesta.
(UE) Uniestructural	Es el nivel en que la respuesta dada a la pregunta atiende a solo una parte de la tarea. Un único proceso o concepto se aplica al menos a un dato. Se extrae una conclusión, pero a menos que el proceso único junto con los datos seleccionados sean suficientes para la correcta solución del problema, la conclusión no será válida del todo.
(ME) Multiestructural	Es el nivel en que la respuesta tiene una descripción cualitativa de la situación. Se utiliza una serie de procesos o conceptos sobre uno o más datos, pero sin síntesis de la información ni conclusiones intermedias. Esta falta de síntesis puede ser aceptable en el caso de una de una pregunta sencilla, o puede indicar un rendimiento cognitivo inferior que se requiere para resolver el problema con éxito.
(R) Relacional	Es el nivel en que la respuesta integra la descripción cualitativa con un aspecto cuantitativo. Una respuesta relacional se caracteriza por la síntesis de información, procesos y resultados intermedios. Para llegar a la conclusión, se aplican conceptos a algunos de los datos, dando resultados intermedios que luego se relacionan con otros datos y/o procesos.
(A+) Abstracto ampliado	Es el nivel más alto en que las respuestas abstractas ampliadas son estructuralmente similares a las respuestas relacionales, pero en este caso los datos, conceptos y/o procesos (más habitualmente los dos últimos) se extraen de fuera del dominio de conocimiento y experiencia que se supone, hipotéticamente, en la pregunta.

Nota. Elaboración propia, a partir de Chick (1998).

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

https://doi.org/10.46219/rechiem.v13i4.80

EXPLORANDO LAS NOCIONES PROBABILÍSTICAS INFORMALES EN ESTUDIANTES DE EDUCACIÓN BÁSICA

EXPLORING INFORMAL PROBABILISTIC NOTIONS IN PRIMARY EDUCATION STUDENTS

Hugo Alvarado Martínez
alvaradomartinez@ucsc.cl
Universidad Católica de la Santísima Concepción, Chile

Sergio Tapia Muñoz
sergiotapia@institutoclaret.cl
Instituto Claret, Temuco, Chile

María Lidia Retamal Pérez
lretamal@ucsc.cl
Universidad Católica de la Santísima Concepción, Chile

Liliana Tauber
estadisticamatematicafhuc@gmail.com
Universidad Nacional del Litoral, Santa Fe, Argentina

RESUMEN

En este trabajo se analizan las intuiciones y heurísticas sobre la probabilidad en 331 estudiantes de primaria (12-13 años) por medio de un cuestionario de seis ítems cerrados, y se examinan las argumentaciones en dos ítems abiertos. Los resultados muestran una alta variación en la intuición probabilística en situaciones aleatorias y la existencia de razonamientos acerca de nociones informales de probabilidad bajo una condición. Se considera pertinente para la enseñanza en este nivel educativo favorecer un acercamiento a la comprensión con distintas situaciones de incertidumbre mediante la profundización de los significados intuitivo y clásico de la probabilidad.

PALABRAS CLAVE:
Probabilidad, Incertidumbre, Intuición, Educación básica.

ABSTRACT

This paper analyzes the intuitions and heuristics about probability in 331 primary school students (12-13 years old), by means of a questionnaire of six closed items and examines the arguments in two open items. The results show a high variation in probabilistic intuition in random situations and the existence of reasoning about informal notions of probability under a condition. It is considered that, at this educational level, it is relevant to favor an approach to understanding with different situations of uncertainty by deepening the intuitive and classical meanings of probability.

KEYWORDS:
Probability, Uncertainty, Intuition, Primary education.

Recibido: 28 de Mayo de 2021, Aceptado: 24 de Junio de 2021

1. Introducción

Un currículo que considera el desarrollo de las intuiciones probabilísticas desde los primeros años, permite enfrentar las dificultades en el aprendizaje de la Estadística y Probabilidad y progresar en el razonamiento formal probabilístico, especialmente confrontando las ideas informales y creencias que tienen los estudiantes sobre las probabilidades (National Council of Teachers of Mathematics, 2000). La enseñanza de la probabilidad en la etapa escolar generalmente no se atiene a las ideas informales y creencias que tienen los alumnos sobre las probabilidades (Kahneman y Tversky, 1972). Más aún, aunque utilizamos nociones probabilísticas informales a diario para tomar decisiones, la investigación sobre probabilidad se ha centrado principalmente en los significados clásico y frecuentista (Alvarado et al., 2018).

Una dimensión de indagación en educación estadística es estudiar las dificultades de comprensión en el razonamiento probabilístico; en particular, investigar cómo las personas hacen juicios y toman decisiones cuando se enfrentan a situaciones de incertidumbre (Garfield y Ben-Zvi, 2008). Se concuerda con Sharma (2014) en que los entornos sociales y la cultura común pueden influir en las ideas informales de probabilidad, y en la necesidad de confrontar las creencias cotidianas intuitivas con los conceptos probabilísticos. Esta confrontación permitiría aclarar los objetivos, el propósito y las limitaciones de la enseñanza de la probabilidad.

Las investigaciones proponen dar oportunidades a los estudiantes de variadas experiencias de situaciones probabilísticas asociadas a los diversos significados de la probabilidad (Batanero et al., 2005). Así, es posible mencionar los estudios relacionados con la enseñanza de la probabilidad mediante paradojas (Batanero et al., 2012; Contreras et al., 2014), probabilidad condicional (Batanero et al., 2014), significados de la probabilidad (English y Watson, 2016; Gómez et al., 2015; Sanabria y Núñez, 2017; Sharma, 2014) y significado de la probabilidad intuitiva (Alvarado et al., 2018; Fulmer, 2014; Gómez et al., 2014).

Teniendo en cuenta estos fundamentos, el presente trabajo analiza las intuiciones y heurísticas sobre la probabilidad que atribuye un grupo de estudiantes de básica (12 y 13 años) en situaciones de incertidumbre, analizando las respuestas entregadas a un cuestionario de seis ítems y las soluciones de dos ítems abiertos. Las interrogantes que rigen este estudio son: ¿Cómo los estudiantes de educación básica responden a situaciones aleatorias que involucran razonamiento heurístico? y ¿Qué argumentos utilizan los estudiantes respecto a nociones de probabilidad clásica?

2. Marco de referencia

2.1 Razonamiento probabilístico

Fischbein (1987) califica como un serio error no considerar la confianza que tienen los estudiantes en sus intuiciones, sugiriendo que hay que tomar conciencia de que se poseen intuiciones correctas y útiles, y que debemos lograr ser capaces de controlar nuestras intuiciones de forma de llegar a comprender (asimilar) de manera adecuada las estructuras formales propias del razonamiento lógico. Fischbein y Schnarch (1997) sugerían que en el aprendizaje de la probabilidad los estudiantes debían crear nuevas intuiciones, y la enseñanza debía proveer de experiencias en que los estudiantes confrontasen sus esquemas intuitivos primarios y las causas de los conflictos y errores en los tipos de razonamiento específicos a las situaciones probabilísticas.

Nisbett y Ross (1980) señalan que es posible adquirir un correcto razonamiento estadístico intuitivo sobre conceptos abstractos, tales como la Ley de los grandes números, y aplicarlo para resolver problemas cotidianos, siempre que reconozcamos la situación como aleatoria. Kahneman et al. (1982) afirman que las heurísticas y sesgos son resistentes a la enseñanza e incluso se observa en sujetos con alta preparación matemática. Estos autores describen los sesgos de razonamiento que ocurren como resultado de un proceso cognitivo, como la heurística, que lleva a una solución inmediata del problema, pero no garantiza que la solución sea correcta, bien por usar un modelo inapropiado de la situación o por falta de estructuras cognitivas específicas. Una heurística puede entenderse como la estrategia utilizada por las personas al emitir un juicio, realizar una estimación, tomar una decisión, entre otras acciones, para descomplejizar un problema, pero basándose en información limitada. Así, a través de heurísticas, las personas reducen la complejidad de calcular probabilidades y predecir valores, pero al no contar con toda la información requerida se producen sesgos de razonamiento en el juicio o decisión que se adopta.

La heurística de la representatividad, como regla intuitiva e informal, permite, a partir de lo que ya se conoce, inferir sobre un suceso, e incluye juzgar la posibilidad de los resultados. En este sentido, un resultado es más probable si su estructura es más similar a la de la asumida por el modelo subyacente. Kahneman y Tversky (1982) indican que, si un suceso es altamente representativo de cierta categoría, se juzga como alta la probabilidad de que el suceso tenga su origen en esa categoría. Sin embargo, su aplicación conduce en muchos casos a inferencias razonables, pues puede aumentar la probabilidad de cometer sesgos debido a que el hecho de que algo sea más representativo no lo hace más probable. En el presente estudio se pretende analizar esta heurística (ítems 1, 2 y 3) con estudiantes de educación básica.

2.2 Significados de la probabilidad

Gómez et al. (2015) indican que la probabilidad, desde su emergencia, ha estado sujeta a diferentes interpretaciones y debates filosóficos que todavía continúan y se relacionan con la concepción y definición del azar en diferentes periodos históricos (Batanero, 2005, 2016; Batanero y Díaz, 2007; Borovcnik y Kapadia, 2014). Batanero (2005) realizó una caracterización de los diferentes significados de la probabilidad y cómo han sido tenidos en cuenta en la enseñanza secundaria. Analiza los diferentes elementos de campos de problemas, procedimientos, lenguaje, propiedades y conceptos relacionados con los cinco significados de la probabilidad. A saber, significado de la probabilidad intuitivo, clásico, frecuencial, subjetivo y axiomático. Este estudio aborda principalmente el significado intuitivo y clásico.

> *a) Significado intuitivo de la probabilidad*: Las ideas intuitivas de probabilidad, como grados de creencia personal, es común encontrarlas al enfrentarse en diversos juegos de azar y también expresadas en noticias y en anuncios publicitarios a través de los medios de comunicación. Batanero (2005) indica que las ideas intuitivas sobre el azar aparecen tanto en niños como en personas que no han estudiado probabilidades, quienes usan frases y expresiones coloquiales como posible, previsible y presumible, para “cuantificar” sucesos inciertos y expresar su grado de creencia en ellos.
>
> b) *Significado clásico de la probabilidad:* Está ligado a la regla de Laplace, donde la probabilidad de un suceso es el cociente entre el número de casos favorables y el número de casos posibles (considerados equiprobables). A pesar de su limitación a experimentos aleatorios con un número finito de posibilidades, este significado ha primado en la escuela durante muchos años en los juegos de azar.

Este trabajo aporta a las investigaciones anteriores (Rodríguez et al., 2018; Tauber y Olesker, 2014; Vásquez y Alsina, 2017) enfocándose en el significado intuitivo (ítems 1, 2, 3 y 6) que apenas ha sido considerado en la enseñanza de la probabilidad de la educación básica, en el significado clásico (ítems 4, 5 y 7) y en la idea informal de probabilidad compuesta (ítem 8). A continuación, se describe la metodología y los resultados obtenidos.

3. Metodología

El procedimiento para llevar a efecto esta investigación considera un estudio descriptivo con el propósito de explorar las nociones probabilísticas informales, para dar respuestas a las interrogantes: ¿Cómo los estudiantes de educación básica responden a situaciones aleatorias que involucran razonamiento heurístico? y ¿Qué argumentos utilizan los estudiantes respecto a nociones de probabilidad clásica? A continuación, se mencionan los participantes, el instrumento utilizado, así como las respuestas esperadas a los ítems propuestos.

3.1 Participantes

Participaron de la investigación 331 estudiantes de segundo ciclo escolar de básica (ver Tabla 1) provenientes de un establecimiento educacional particular gratuito de una región del sur de Chile. De los 331 estudiantes, de edades entre 12-13 años, 172 (51,2%) son mujeres. Los estudiantes completaron el cuestionario con lápiz y papel en una sesión de dos horas pedagógicas (total 90 minutos) en la asignatura de Matemática. Cabe señalar que los estudiantes de este nivel han realizado conjeturas acerca de la tendencia de resultados obtenidos en repeticiones de un mismo experimento con dados, monedas u otros, y explican las probabilidades de eventos obtenidos por medio de experimentos de manera manual, estimándolas de manera intuitiva, utilizando frecuencias relativas y relacionándolas con razones, fracciones o porcentaje.

Tabla 1. Número de estudiantes según nivel educativo

Nota. Elaboración propia.

Nivel	Número de estudiantes (%)
Séptimo	156 (47,1%)
Octavo	175 (52,9%)

3.2 Instrumento

Se construyó un cuestionario con ocho ítems, seis de ellos cerrados y dos abiertos. Los ítems evaluaban intuiciones, heurísticas y conocimientos básicos de probabilidad en estudiantes de educación básica. Los ítems 1, 2 y 3 del cuestionario presentan heurísticas de representatividad y fueron adaptados de los problemas de Kahneman y Tversky (1972) y Tversky y Kahneman (1974). El ítem 6 es una adaptación del problema de Gardner (1959) y hace mención a la confusión de probabilidades del producto y condicional. En este ítem se les propuso a los participantes evaluar en escala ordinal de 10 en 10 para estimar su grado de creencia sobre probabilidades, dentro de un rango del 0 al 100. El ítem 4 aplica la regla de Laplace en extracciones sin reemplazo y el ítem 5 estudia la regla de Laplace al aumentar el número de repeticiones. El ítem 7 fue seleccionado de Vásquez y Alsina (2017), donde compara extracciones de bolitas por la regla de Laplace. El ítem 8 se refiere a la aplicación de la regla del producto y fue elaborado por los autores.

Un cuestionario más amplio, que incluye estos ítems, fue evaluado por cuatro especialistas de estadística y didáctica de la estadística y aplicado a estudiantes de otros niveles educativos (Alvarado et al., 2018). Cabe señalar que este estudio pretende explorar las intuiciones probabilísticas y aplicación de la probabilidad clásica de los estudiantes de nivel escolar sobre estos ítems, dado que posteriormente, en el nivel medio-superior, son formalizadas sus soluciones.

En la Tabla 2 se describen los ocho ítems de cuestionario aplicado y su objetivo de evaluación, presentando situaciones referidas a nociones probabilísticas informales con ítems de selección múltiple y la estimación de probabilidad en escala ordinal con apreciación de 0 a 100 (ítem 6).

Tabla 2 . Descripción de los ítems de razonamiento y conocimientos de probabilidad en el cuestionario

Nota. Elaboración propia.

Ítem	**Razonamiento emergente**
1. Si de una tómbola con pelotas enumeradas del 1 al 5 se sacan cinco pelotas con reposición, ¿qué es más probable que ocurra? a) Que salga el número 22222. b) Que salga el número 12345. c) Que salga el número 25314. d) Todas son igualmente probables.	Falacia del jugador.
2. Se lanza una moneda 8 veces, obteniendo en orden los siguientes resultados: cara, sello, cara, sello, sello, sello, sello, sello. Si se lanza la moneda por novena vez, ¿qué es más probable que pase? a) Es más probable que salga cara, puesto que han salido demasiados sellos y ya es hora de que salga cara. b) Es más probable que salga sello, puesto que ha salido sello en cinco lanzamientos sucesivos. c) Es igual de probable que salga cara o sello.	Heurística de representatividad. Independencia de eventos.
3. Eduardo desde pequeño mostró gran afición por el arte. Entró a estudiar a la universidad donde se destacó por su talento por la escritura y fotografía, transformándose en el mejor de su carrera. En la última década recorrió el mundo fotografiando y ayudando en los diversos desastres naturales. ¿Cuál de estos sucesos tiene más probabilidad de ser cierto? a) Eduardo trabaja en un diario. b) Eduardo trabaja en un diario y es voluntario de bomberos. c) Ambos sucesos son igualmente probables.	Heurística de representatividad.
4. En una bolsa se ponen 4 bolas rojas, 4 azules y 2 verdes, y después se mezclan. Se sacan 3 bolas fuera, resultando 2 rojas y 1 azul. A continuación, se saca otra bola sin echar las anteriores. ¿De qué color es más probable que sea? a) El rojo tiene mayor probabilidad. b) El azul tiene mayor probabilidad. c) El verde tiene mayor probabilidad.wd) Todos los colores tienen la misma probabilidad.	Razonamiento proporcional. Regla de Laplace.
5. Cindy y Trudy juegan tirando un dado normal. Si sale un 5 gana Cindy y si sale menos de 3 gana Trudy. ¿Cuántas veces habrá ganado cada uno, aproximadamente, después de tirar el dado 60 veces? a) Trudy y Cindy ganan el mismo número de veces. b) Cindy gana el doble de veces que Trudy. c) Trudy gana el doble de veces que Cindy. d) Cindy gana 17 veces más que Trudy. e) Trudy gana 17 veces más que Cindy.	Razonamiento proporcional. Regla de Laplace.
6. Si se lanza una moneda tres veces, ¿qué tan probable es obtener sello en el tercer lanzamiento si se sabe que los dos primeros lanzamientos fueron cara? Asignar valores de posibilidad de 0 a 100.	Independencia de eventos. Probabilidad bajo condición.

<table>
<tr><td>7. Resuelva. Eduardo tiene en su caja 10 bolas blancas y 20 negras. Luis tiene en su caja 30 bolas blancas y 60 negras. Juegan una partida de azar. El ganador es el niño que saque primero una bola blanca. Si ambos sacan simultáneamente una bola blanca o una bola negra, ninguno gana, devuelven las bolas a las cajas y la partida continúa. Eduardo afirma que el juego no es justo porque en la caja de Luis hay más bolas blancas que en la suya.
a) El juego es justo.
b) Luis tiene más ventaja.
c) Eduardo tiene más ventaja.</td><td>Razonamiento proporcional. Regla de Laplace.</td></tr>
<tr><td>8. Resuelva. Se muestran los resultados de una encuesta realizada a 60 personas sobre la preferencia de mermeladas, clasificadas en no dietética y dietética. Al seleccionar a uno de estos encuestados al azar, la probabilidad de que sea hombre y prefiera mermelada no dietética es:
<table><tr><th></th><th>Mermelada no dietética</th><th>Mermelada dietética</th></tr><tr><td>Mujer</td><td>6</td><td>24</td></tr><tr><td>Hombre</td><td>18</td><td>12</td></tr></table>
a) 18/24
b) 18/30
c) 18/60
d) 30/60 × 24/60
e) 1/18</td><td>Regla del producto.</td></tr>
</table>

3.3 Respuestas esperadas a los ítems cerrados

El enunciado del ítem 1, asociado a los juegos de azar, es conocido como la falacia del jugador o falacia de Montecarlo. Esto es, al ser una falacia lógica, se cree incorrectamente que los sucesos ya ocurridos afectan a los que podrían ocurrir. La respuesta correcta es la opción (d), debido a que la probabilidad de la secuencia 22222 es de un 0,032%, la probabilidad de la secuencia 12345 es de un 0,032% y la probabilidad de la secuencia 25314 es de un 0,032%, por lo tanto, todas las opciones tienen la misma posibilidad de ocurrencia. Tversky y Kahneman (1974, p. 1125) han indagado este tipo de preguntas, llegando a la conclusión de "que los individuos consideran que un suceso aleatorio tiene menos posibilidad de ocurrir porque ha ocurrido durante cierto periodo", en este caso el valor 2 de la secuencia.

El ítem 2 requiere identificar la presencia de un evento independiente, debido a que solamente se trabajará con una sola moneda. Así, los eventos resultantes pueden ser cara (C) o sello (S), teniendo cada una ½ de probabilidad de ocurrir. Al ser solo una moneda, los resultados anteriores no tienen incidencia en el próximo lanzamiento, ya que la posibilidad de ocurrencia no se altera ni varía. Por lo tanto, la respuesta correcta corresponde a la alternativa (c). Situaciones de este tipo han sido estudiadas por Serrano et al. (1998) y Tauber y Olesker (2014). En sus investigaciones sobre posibilidad de ocurrencia de 5 monedas con 5 opciones dirigido a estudiantes de educación media, evalúan si los estudiantes usan la heurística de la representatividad en sus juicios sobre la probabilidad de obtener diferentes secuencias en el lanzamiento de una moneda.

El ítem 3 es una adaptación de Tversky y Kahneman (1974). Hay que tener en cuenta las probabilidades de dos situaciones: a) "Eduardo trabaja en un diario" b) Eduardo trabaja en un diario y es voluntario de bomberos"; para que las ocurrencias de dos sucesos sean simultáneas, deben ser menor o igual que la probabilidad de ocurrencia de cada uno. La solución a este problema es la alternativa (a) en donde $P(A \cap B) \leq P(A)$ y $P(A \cap B) \leq P(B)$. Tversky y Kahneman (1974) señalan que la mayoría de quienes responden esta problemática utilizan la heurística de representatividad en este tipo de pregunta, dejando de lado las propiedades de la probabilidad.

El ítem 4 es un problema de asignación de probabilidades en extracción sin reemplazo. La solución requiere de razonamiento proporcional, al comparar tres posibles sucesos que no son de igual probabilidad con extracciones sin reemplazo, lo que modifica la composición de la urna y el espacio muestral. En este caso es importante el suceso previo y su dependencia en el siguiente suceso. La respuesta correcta es (b).

El ítem 5 tiene como respuesta correcta la alternativa (c). Un dado normal tiene 6 caras y la probabilidad de que salga 5 es de 1/6 por parte de Cindy. Al ser 60 lanzamientos, se debe multiplicar 1/6 por 60, dando como producto 10. En el caso de Trudy, gana cuando el dado es menor que 3 (1 ó 2), en donde la probabilidad de obtener esos números es 2/6; al ser lanzado 60 veces, se debe multiplicar 2/6 por 60, lo que da como resultado 20.

En el ítem 6 hay que tener presente que el resultado (cara o sello) es independiente, ya que solo se trabaja con una moneda, por lo que la probabilidad de que sea sello es de ½; aunque los lanzamientos anteriores sean cara, la probabilidad de ocurrencia sigue siendo la misma.

En el ítem 7, aunque Eduardo tiene menor cantidad de bolas blancas en comparación con las negras, la probabilidad de extraer una de color blanco es de 1/3, ya que son 10 blancas dentro de un total de 30. Las opciones que tiene Luis de sacar la bola blanca son de 30 dentro de un total de 90; aunque tiene más bolas blancas, la probabilidad de extracción, al igual que Eduardo, corresponde a 1/3, lo que conlleva a que el juego sea justo (respuesta correcta (a)). Este ítem corresponde a un estudio de Vásquez y Alsina (2017), en el que analizaron los resultados con profesores que enseñan Matemática.

En el ítem 8 los estudiantes deben aplicar los conocimientos relacionados con la probabilidad del producto o regla de la multiplicación, en este caso que sea hombre (30) y que prefiera la mermelada no dietética (18). La solución correcta, alternativa (c), tiene por probabilidad 18/60.

4. Resultados

4.1 Resultados de los ítems cerrados 1 al 6 según asignaciones de probabilidad intuitiva y clásica

A continuación, se realiza un análisis descriptivo de los ítems que involucra razonamiento de la heurística de representatividad (ítems 1, 2, 3 y 6), conocimiento de la probabilidad clásica (ítems 4, 5 y 7) y conocimiento de la regla del producto (ítem 8), respondidos por 331 estudiantes de educación básica.

Ítem 1. Si de una tómbola con pelotas enumeradas del 1 al 5 se sacan cinco pelotas con reposición, ¿qué es más probable que ocurra?
a) Que salga el número 22222.
b) Que salga el número 12345.
c) Que salga el número 25314.
d) Todas son igualmente probables.

En la Figura 1 se puede apreciar que 69,7% de los estudiantes seleccionan la alternativa (d), que corresponde a la respuesta correcta, mientras que 27,3% de los participantes optan por las alternativas (a), (b) y (c). Las investigaciones de Tversky y Kahneman (1974) indican como error frecuente que los estudiantes crean que los sucesos pasados pueden afectar en los que ya vienen en lo relativo a actividades aleatorias, suponiendo más probable que salgan números en orden aleatorio (alternativa c) que números ordenados o repetidos (alternativas a y b).

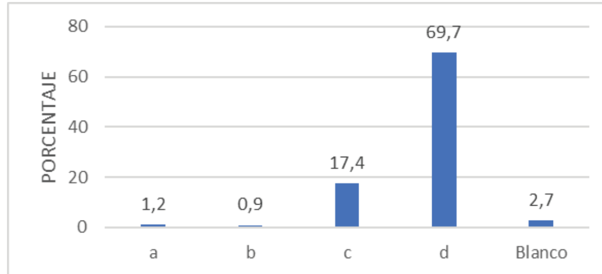


Figura 1. Porcentaje de respuestas al ítem 1, n = 331 (respuesta esperada d). Nota. Elaboración propia.

Ítem 2. Se lanza una moneda 8 veces, obteniendo en orden los siguientes resultados: cara, sello, cara, sello, sello, sello, sello, sello. Si se lanza la moneda por novena vez, ¿qué es más probable que pase?

a) Es más probable que salga cara, puesto que han salido demasiados sellos y ya es hora de que salga cara.

b) Es más probable que salga sello, puesto que ha salido sello en cinco lanzamientos sucesivos.

c) Es igual de probable que salga cara o sello.

En la Figura 2 se observa que 77,9% de los estudiantes seleccionaron la alternativa correcta (c), señalando que existe la misma probabilidad de que salga cara o sello, sin considerar los resultados de los lanzamientos anteriores. Además, es posible visualizar que el 20,9% de los estudiantes cree que sí inciden los sucesos anteriores en el noveno lanzamiento (alternativas a y b), situación similar a la presentada en el ítem 1. Si se comparan estos resultados con los obtenidos en la investigación de Vásquez y Alsina (2017), en los que un 36,6% de los profesores de educación básica en ejercicio que enseñan Matemática dieron una respuesta correcta, se verá que en este estudio se obtuvieron mejores resultados.

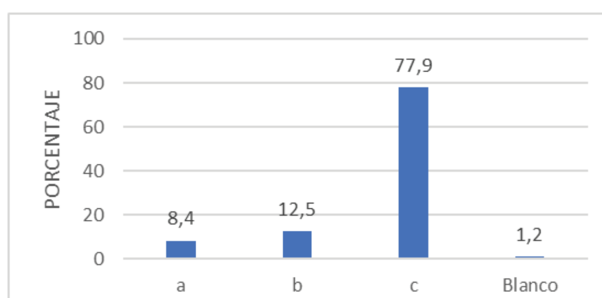


Figura 2. Porcentaje de respuestas al ítem 2, n = 331 (respuesta esperada c). Nota. Elaboración propia.

Ítem 3. Eduardo desde pequeño mostró gran afición por el arte. Entró a estudiar a la universidad donde se destacó por su talento por la escritura y fotografía, transformándose en el mejor de su carrera. En la última década recorrió el mundo fotografiando y ayudando en los diversos desastres naturales. ¿Cuál de estos sucesos tiene más probabilidad de ser cierto?

a) Eduardo trabaja en un diario.
b) Eduardo trabaja en un diario y es voluntario de bomberos.
c) Ambos sucesos son igualmente probables.

En la Figura 3 es posible observar que el 35,3% de los estudiantes ha seleccionado la alternativa (a), el 23,9% considera correcta la alternativa (b) y el 39,9% cree correcta la alternativa (c). Estos resultados discrepan de los obtenidos por Tversky y Kahneman (1974), encontrando que la mayoría de los encuestados escoge la alternativa (b) influenciándose por la heurística de la representatividad en este tipo de preguntas, dejando de lado las propiedades de la probabilidad.

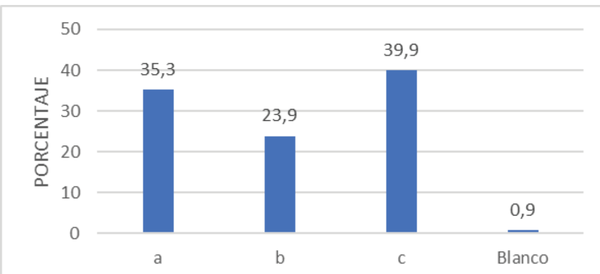


Figura 3. Porcentaje de respuestas al ítem 3, n = 331 (respuesta esperada a). Nota. Elaboración propia.

Ítem 4. En una bolsa se ponen 4 bolas rojas, 4 azules y 2 verdes, y después se mezclan. Se sacan 3 bolas fuera, resultando 2 rojas y 1 azul. A continuación, se saca otra bola sin echar las anteriores. ¿De qué color es más probable que sea?

a) El rojo tiene mayor probabilidad.
b) El azul tiene mayor probabilidad.
c) El verde tiene mayor probabilidad.
d) Todos los colores tienen la misma probabilidad.

En la Figura 4 se aprecia que 52,5% de los estudiantes señalaron que la bola de color azul tiene mayor probabilidad de salir en una segunda extracción sin reemplazo (opción b correcta), resultados similares a los obtenidos por Cañizares (1997) en donde un 50% de los futuros profesores de Matemática testeados respondió correctamente. Los distractores en el problema de razonamiento proporcional asignando probabilidades en extracciones sin reemplazo provocó que 26,1% de los alumnos escogiera la alternativa (d), ya que es la opción de equiprobabilidad, entendiendo así que al ser bastante baja la diferencia de las bolas de los distintos colores que hay dentro de la urna, no existe mayor influencia al haber mayor o menor probabilidad de extraer una de las bolas. En otro caso con profesores de Matemática, 91,5% respondió correctamente (Bastías, 2017), teniendo así una diferencia considerable respecto de los resultados de este estudio.

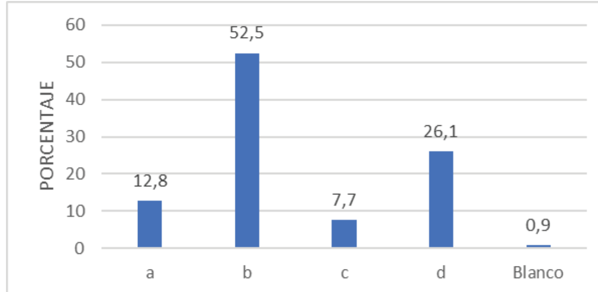


Figura 4. Porcentaje de respuestas al ítem 4, n = 331 (respuesta esperada b). Nota. Elaboración propia.

Ítem 5. Cindy y Trudy juegan tirando un dado normal. Si sale un 5 gana Cindy y si sale menos de 3 gana Trudy. ¿Cuántas veces habrá ganado cada uno, aproximadamente, después de tirar el dado 60 veces?

a) Trudy y Cindy ganan el mismo número de veces.
b) Cindy gana el doble de veces que Trudy.
c) Trudy gana el doble de veces que Cindy.
d) Cindy gana 17 veces más que Trudy.
e) Trudy gana 17 veces más que Cindy.

En la Figura 5 se puede observar que 36,8% de los estudiantes señala correctamente que Trudy gana el doble de veces que Cindy (alternativa c). Los distractores fueron similares en porcentaje de respuestas, 18,2% de los estudiantes indicaron la alternativa (a), que Trudy y Cindy ganan el mismo número de veces, seguido de 17,7% que seleccionaron que Trudy gana 17 veces más que Cindy (alternativa e). Los resultados de Bastías (2017) fueron mejores con profesores de Matemática, alcanzando 79,9% de respuestas correctas, lo que permite dilucidar que tener la formación matemática otorga el uso de propiedades tales como la regla de Laplace y la esperanza matemática.

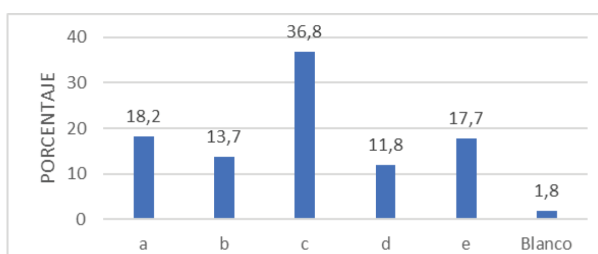


Figura 5. Porcentaje de respuestas al ítem 5, n = 331 (respuesta esperada c). Nota. Elaboración propia.

Ítem 6. Si se lanza una moneda tres veces, ¿qué tan probable es obtener sello en el tercer lanzamiento si se sabe que los dos primeros lanzamientos fueron cara?

En la Figura 6 se puede apreciar que 49,2% de los estudiantes asignó un valor de 50 de probabilidad de obtener sello en el tercer lanzamiento, respuesta que es correcta, debido a que las opciones que se presentan son independientes entre sí (cara y sello). Además, un 33,6% otorgó un valor inferior a 50, destacando que 12% declaró valores entre 10 y 20 de probabilidad de ocurrencia, y que puede ser explicado al pensar el suceso CCS como uno de los ocho sucesos del espacio muestral sin atender a la condicionalidad del problema. Estos resultados se condicen con la experiencia similar en niños de 11 a 13 años de edad (Estrella et al., 2019), señalando que

> Cada uno de los dos posibles resultados, cara o sello, tienen la misma probabilidad independientemente del número de veces que la moneda se haya lanzado antes, y de los resultados obtenidos. Razonar que es más probable que el próximo lanzamiento será sello en vez de cara debido a los anteriores lanzamientos es la falacia: la idea de que una racha de suerte pasada influya de alguna forma en las posibilidades futuras. (p. 294)

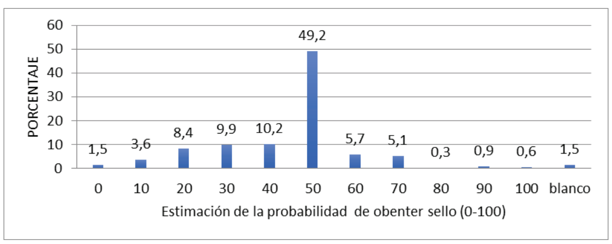


Figura 6. Porcentaje de respuestas al ítem 6 según estimación de la probabilidad (respuesta esperada valor 50), n = 331
Nota. Elaboración propia.

4.2 Resultados de los ítems cerrados 1 al 6 según nivel educativo

En la Figura 7 se presenta la distribución de las respuestas correctas de los primeros seis ítems cerrados dadas por los 331 estudiantes de educación básica. El gráfico permite comparar la variabilidad de cada ítem, observando que las diferencias entre los dos niveles son mínimas, siendo los ítems 1 y 3 sobre heurística de representatividad los de mayores diferencias con 8,1% y 8,9%, respectivamente. Por otra parte, el ítem 6 es el que ha presentado menores dificultades en ambos grupos, resultado importante a la hora de reflexionar sobre el razonamiento intuitivo de los estudiantes, ya que como se señaló anteriormente, la actividad conduce a pensar sobre probabilidades condicionales e independencia. Lo anterior, puede ser muy relevante en el sentido de desarrollar propuestas didácticas que potencien estos razonamientos; como es el caso de la falacia del jugador y la heurística de la representatividad, que según Fischbein (1975), los estudiantes carecen de experiencias en ambientes de desarrollo de intuiciones primarias sobre la independencia de sucesos.

Figura 7. Variabilidad de respuestas correctas, ítems 1 a 6 según nivel educativo. Nota. Elaboración propia.

4.3 Análisis de los argumentos a las respuestas de los ítems 7 y 8

Ítem 7. Argumenta. Eduardo tiene en su caja 10 bolas blancas y 20 negras. Luis tiene en su caja 30 bolas blancas y 60 negras. Juegan una partida de azar. El ganador es el niño que saque primero una bola blanca. Si ambos sacan simultáneamente una bola blanca o una bola negra, ninguno gana, devuelven las bolas a las cajas y la partida continúa. Eduardo afirma que el juego no es justo porque en la caja de Luis hay más bolas blancas que en la suya.

a) El juego es justo.
b) Luis tiene más ventaja.
c) Eduardo tiene más ventaja.

El 35,3% de los estudiantes responde correctamente, señalando que el juego es justo entre Luis y Eduardo (ver Figura 8. Si bien el espacio muestral de cada uno de ellos es distinto en cantidad, no lo es en proporcionalidad, por lo tanto, existe la misma probabilidad de 1/3 en obtener una bola blanca (opción a).

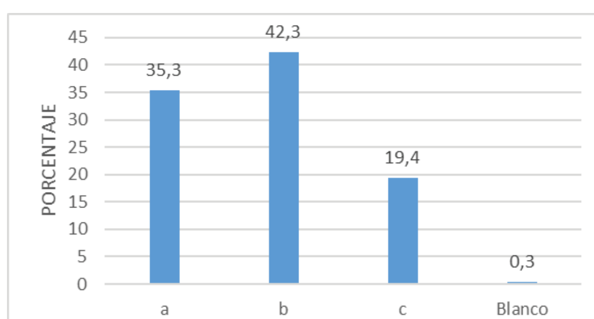


Figura 8. Porcentaje de respuestas al ítem 7, n = 331
Nota. Elaboración propia.

Hubo respuestas que si bien aplicaron la proporcionalidad no consideran el espacio muestral adecuado, es decir, en el caso de Eduardo consideraron 10/20 y para Luis 30/60 (ver Figura 9). Cabe destacar, en este caso, avances en la conexión de la aritmética con la probabilidad, es decir, los estudiantes están relacionando los conceptos matemáticos de porcentaje, razón, proporcionalidad directa y probabilidad.

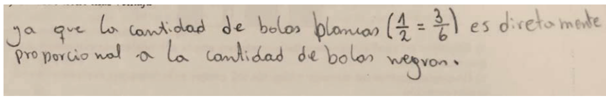


Figura 9. Argumento correcto del ítem 7

Nota. Elaboración propia.

Al comparar los resultados con la investigación realizada por Vásquez y Alsina (2017) con profesores de educación básica que enseñan Matemática, estos obtuvieron una mayor cantidad de respuestas correctas, con un 60,2%. En el caso de este estudio, 42,2% de los estudiantes optaron por que Luis tiene más ventaja, alternativa (b), asumiendo que al existir un espacio muestral más grande (90 bolas en total) es más difícil obtener la bola blanca, tal como lo argumenta un estudiante (ver Figura 10).

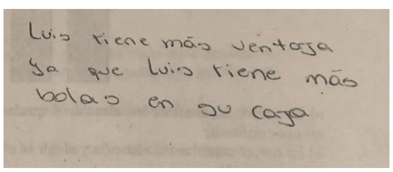


Figura 10. Argumento incorrecto del ítem 7

Nota. Elaboración propia.

También, 19,3% de los estudiantes señala que Eduardo tiene más posibilidades de ganar, siendo esta la alternativa (c). Una respuesta es dada en la Figura 11.

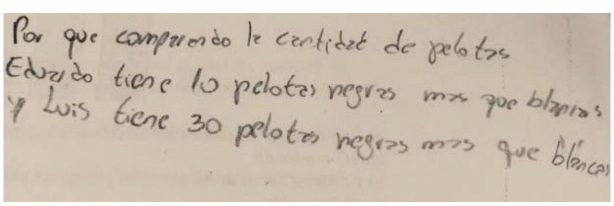


Figura 11. Argumento incorrecto del ítem 7

Nota. Elaboración propia.

La Tabla 3 no muestra diferencias importantes según nivel educativo de los estudiantes, tanto para la respuesta correcta e incorrectas, solo hay 2,7% de diferencia entre los alumnos de séptimo y octavo en su respuesta correcta.

Tabla 3 . Resultados del ítem 7 según nivel educativo en porcentaje

Nota. Elaboración propia.

	Séptimo	**Octavo**	**Total**
a	16,3	19,0	35,3
b	21,5	20,8	42,3
c	7,9	11,5	19,4
Blanco	1,5	1,5	3,0

Ítem 8. Argumenta. Se muestran los resultados de una encuesta realizada a 60 personas sobre la preferencia de mermeladas, clasificadas en no dietética y dietética.

	Mermelada no dietética	**Mermelada dietética**
Mujer	6	24
Hombre	18	12

Al seleccionar a uno de estos encuestados al azar, la probabilidad de que sea hombre y prefiera mermelada no dietética es:

a) 18/24 b) 18/30 c) 18/60
d) 30/60 × 24/60 e) 1/18

Del total de estudiantes hubo un 37,1% de respuestas correctas, opción c (ver Figura 12).

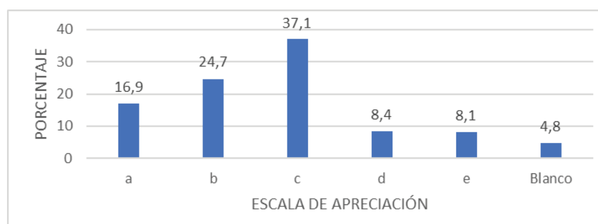


Figura 12. Porcentaje de respuestas al ítem 8, n = 331

Nota. Elaboración propia.

La Figura 13 presenta una solución correcta de obtener una probabilidad 18/60 en seleccionar a una persona hombre que prefiera mermelada no dietética, utilizando un razonamiento mediante la regla de Laplace.

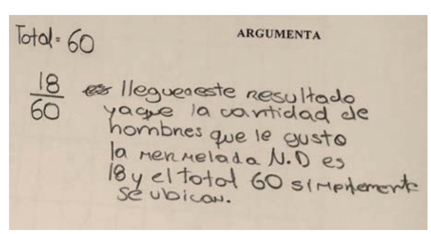


Figura 13. Argumento correcto del ítem 8

Nota. Elaboración propia.

En cambio, 24,4% de los estudiantes (Figura 14) asumen como correcta la alternativa (b) 18/30, considerando el espacio muestral de hombres (30) y dejando de lado a la cantidad de mujeres que indica el problema. Esta opción fue la más alta de error, y es interesante puesto que están dando una respuesta de probabilidad condicionada al reducir el espacio muestral, error que ha sido señalado por Contreras (2011) respecto a futuros profesores de Matemática, al confundir la probabilidad del producto y probabilidad condicional. Un estudiante señaló lo siguiente (ver Figura 14):

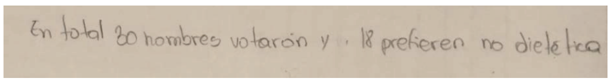


Figura 14. Argumento incorrecto del ítem 8
Nota. Elaboración propia.

En tanto, el 16,9% de los estudiantes (Figura 15) señaló la alternativa (a) de probabilidad 18/24, considerando el espacio muestral reducido de personas que prefieren mermelada no dietética, y de ellos calcularon la posibilidad de seleccionar a una persona hombre, pensada como probabilidad condicionada. Ver ejemplo en la Figura 15.

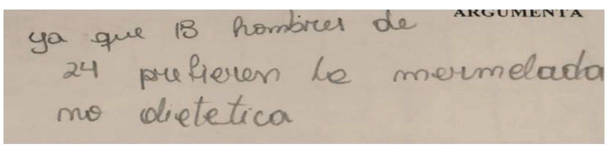


Figura 15. Argumento incorrecto del ítem 8
Nota. Elaboración propia.

La alternativa (d) de probabilidad 30/60×24/60 fue elegida por el 8,4% de los encuestados, en donde los estudiantes han argumentado de la siguiente manera la elección de su respuesta (Figura 16).

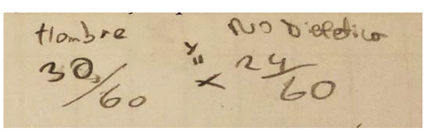


Figura 16. Argumento incorrecto del ítem 8
Nota. Elaboración propia.

Por último, hubo otro argumento que no está en contexto matemático, pero sí lo está en los temas que conciernen a las áreas de la salud: por lo general, las personas que consumen productos dietéticos son quienes sufren de alguna enfermedad, como lo que se puede apreciar en la siguiente Figura 17.

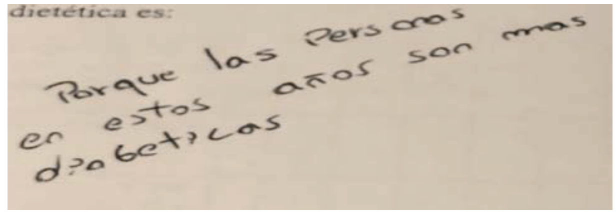


Figura 17. Argumento incorrecto del ítem 8
Nota. Elaboración propia.

En la Tabla 4 se puede apreciar que existe una diferencia de un 3,3% en la alternativa correcta de los estudiantes de séptimo y octavo año de educación básica. Además, no se observan diferencias importantes en los porcentajes de los distractores de acuerdo al nivel que se encuentra el estudiante.

Tabla 4 . Resultados del ítem 8 según nivel educativo en porcentaje
Nota. Elaboración propia.

	Séptimo	**Octavo**	**Total**
a	8,8	8,1	16,9
b	11,2	13,5	24,7
c	16,9	20,2	37,1
d	3,0	5,4	8,4
e	4,2	3,9	8,1
Blanco	3,0	1,8	4,8

5. Conclusiones

Este estudio hace mención a la importancia de tener en cuenta los significados intuitivo y clásico de la probabilidad a temprana edad, por el alcance que pueden tener en la construcción y modelación de la probabilidad frecuencial, subjetiva y axiomática en los niveles de secundaria y superior. En específico, se analizan los significados intuitivo y clásico de la probabilidad que atribuyen 331 estudiantes de educación básica de los niveles séptimo y octavo (12-13 años), por medio de las respuestas de los participantes a un cuestionario de seis ítems cerrados y analizando las argumentaciones de dos ítems abiertos.

En relación a la primera pregunta de investigación: ¿Cómo los estudiantes de educación básica responden a situaciones aleatorias que involucran razonamiento heurístico?, los resultados del cuestionario relacionados con la heurística de representatividad y la independencia de eventos revelan que en los ítems 1 y 2 no hubo mayor dificultad en las situaciones con experimentos de selección de pelotas con reposición y lanzamiento de monedas, obteniendo respuestas correctas sobre 70%, poniendo en juego la equiprobabilidad. El ítem 6, referido a la posibilidad de obtener en el tercer lanzamiento sello sabiendo que hubo dos caras, obtuvo tan solo un 49% de aciertos, lo que puede ser explicado por que hubo estudiantes que asociaron el experimento con el espacio muestral compuesto por ocho elementos, de los cuales hay un caso posible {ccs}. En el ítem 3, presentado a los estudiantes como una situación novedosa de la vida cotidiana (adaptado de Tversky y Kahneman, 1974), 35% alcanzaron la respuesta adecuada siendo la de mayor dificultad puesto que debían considerar dos situaciones, la ocurrencia de un suceso y dos sucesos simultáneamente. Las investigaciones sobre razonamiento probabilístico señalan que ante preguntas de respuesta simple se produce un sesgo predecible de nuestra mente, donde la persona contesta obviando ciertos datos estadísticos. En particular, en este estudio se ha ampliado el campo de aplicación de las intuiciones y heurísticas sobre la probabilidad en este nivel educativo (ítem 3), limitado principalmente a los sorteos y juegos de azar presentes en la probabilidad informal.

Las orientaciones curriculares de Matemática señalan que "en el área de la Probabilidad se pretende que estimen de manera intuitiva y que calculen de manera precisa la probabilidad de ocurrencia de eventos" (Ministerio de Educación de Chile, 2015). Sin embargo, en este nivel educativo son escasas las propuestas, fuera de los dispositivos manipulativos de fichas, monedas y dados, con planteamiento de situaciones sencillas y reales de donde emerjan razonamientos de heurística y presencia de nociones elementales de probabilidad.

En relación a los ítems que incluyeron nociones probabilísticas informales, se notó que en situaciones de razonamiento proporcional y regla de Laplace (ítems 4 y 5) hubo confusiones de los estudiantes al considerar de forma incorrecta los experimentos como equiprobables, 26,1% y 18,2%, respectivamente. A pesar, que los profesores de educación básica promueven la elaboración, con material concreto de dados y monedas, experimentos aleatorios con resultados equiprobables y no equiprobables.

En el análisis de las respuestas a los distintos ítems 1 a 6, se observaron diferencias mínimas según el nivel que cursaban los estudiantes, existiendo diferencias de tan solo 8,1% y 8,9% en los ítems 1 y 3 sobre heurística de representatividad (Figura 7). Así, la variabilidad de las respuestas no es alta entre un nivel y otro. Cabe señalar que estos dos niveles consecutivos son de transición a la educación media.

En las argumentaciones que conllevan elementos de la aritmética (ítem 7), los estudiantes situaron los conceptos de proporcionalidad directa, porcentaje, razón y probabilidad. Llama la atención que 42,2% de los estudiantes respondieron como distractor, mediante la aplicación de la proporcionalidad, pero no consideraron el espacio muestral adecuado. Continuando con la segunda pregunta de investigación: ¿Qué argumentos utilizan los estudiantes respecto de nociones de probabilidad clásica?, el ítem 8 resultó interesante, pues si bien la lectura de tablas de doble entrada se promueve en los siguientes niveles del currículo de Matemática, se intentó indagar en las argumentaciones de nociones de probabilidad informales como la regla del producto. 37,1% utilizaron elementos de probabilidad compuesta, estimando la probabilidad de la ocurrencia de dos sucesos simultáneamente. Nótese en las argumentaciones de los estudiantes respuestas de nociones de probabilidad bajo una condición, declarando una reducción del espacio muestral en su razonamiento condicionado informal, siendo esta la opción de 41,3% de los 331 estudiantes participantes. La literatura también reporta confusiones entre la regla del producto y la probabilidad condicional en profesores de Matemática.

Debido a los escasos estudios sobre el tema en este ciclo educativo, al parecer los estudiantes de educación básica no han sido confrontados a una enseñanza de la probabilidad que evalúe y realce las intuiciones probabilísticas. Los resultados obtenidos del cuestionario sobre elementos de probabilidad informal pueden ser útiles para que los profesores promuevan la estimación intuitiva de la probabilidad. En base a la experiencia docente de los autores del presente estudio, estos coinciden con Fischbein (1975) en que los estudiantes, si bien no tienen un pensamiento formal sobre el azar, pueden desarrollar intuiciones relacionadas a este, lo que favorecería la comprensión de los conceptos probabilísticos. Se considera pertinente elaborar diseños de enseñanza de la probabilidad para estudiantes de educación básica que tengan presente no solo el trabajo con material concreto de azar, sino también ampliar con actividades de estimación de intuiciones probabilísticas de contingencia e interés de los estudiantes.

Es deseable continuar con la investigación, considerando una muestra más amplia de estudiantes que tenga en cuenta otros niveles escolares y distintos establecimientos educacionales (municipalizados y particular pagados). Así también, implementar y evaluar ciclos formativos de razonamientos informales de probabilidad. Una limitación del estudio ha sido confrontar algunos de los resultados con experiencias similares de profesores que enseñan probabilidad, debido a que no hay indagaciones relacionadas con este estudio de asignación de probabilidad intuitiva y exploración de ideas informales de probabilidad compuesta y probabilidad condicional.

Referencias

Alvarado, H., Estrella, S., Retamal, L., y Galindo, M. (2018). Intuiciones probabilísticas en estudiantes de ingeniería: implicaciones para la enseñanza de la probabilidad. *Revista Latinoamericana de Investigación en Matemática Educativa, 21*(2), 131-156. https://doi.org/10.12802/relime.18.2121

Bastías, H. (2017). E*studio del significado intuitivo y formal de la probabilidad en profesores de matemática* [Tesis de Magister no publicada]. Universidad Católica de la Santísima Concepción.

Batanero, C. (2005). Significados de la probabilidad en la educación secundaria. *Revista Latinoamericana de Matemática Educativa, 8*(3), 247-264.

Batanero, C. (2016). Understanding randomness: Challenges for research and teaching. En K. Krainer y N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 34-49). European Society for Research in Mathematics Education.

Batanero, C., Contreras, J. M., Cañadas, C., y Gea, M. M. (2012). Valor de las paradojas en la enseñanza de las matemáticas. Un ejemplo de probabilidad. *Novedades educativas, 261*, 78-84.

Batanero, C., Contreras, J. M., y Díaz, C. (2014). Sesgos en el razonamiento sobre probabilidad condicional e implicaciones para la enseñanza. *Revista Digital: Matemática, Educación e Internet, 12*(2), 1-13. https://doi.org/10.18845/rdmei.v12i2.1673

Batanero, C., y Díaz, C. (2007, julio). *Probabilidad, grado de creencia y proceso de aprendizaje* [Ponencia invitada]. XIII Jornadas Nacionales de Enseñanza y Aprendizaje de las Matemáticas. Federación Española de Profesores de Enseñanza de las Matemáticas, Granada, España.

Batanero, C., Henry, M., y Parzysz, B. (2005). The nature of chance and probability. En G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 16-42). Springer.

Borovcnik, M., y Kapadia, R. (2014). A historical and philosophical perspective on probability. En E. J. Chernoff y B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 7-34). Springer. https://doi.org/10.1007/978-94-007-7155-0_2

Cañizares, M. J. (1997). *Influencia del razonamiento proporcional y combinatorio y de creencias subjetivas en las intuiciones probabilísticas primarias* [Tesis Doctoral, Universidad de Granada]. Repositorio UGR. http://www.ugr.es/~batanero/pages/ARTICULOS/CANIZARE.pdf

Contreras, J. M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional.* [Tesis Doctoral, Universidad de Granada]. Repositorio UGR. http://hera.ugr.es/tesisugr/19831870.pdf

Contreras, J. M., Batanero, C., Arteaga, P., y Cañadas, G. (2014). La paradoja del niño o niña: aplicaciones para la clase de probabilidad. *Revista digital Matemática, Educación e Internet, 14*(1), 1-13. https://doi.org/10.18845/rdmei.v14i1.1562

English, L., y Watson, J. (2016). Development of probabilistic understanding in fourth grade. Journal for *Research in Mathematics Education, 47*(1), 28-62. https://doi.org/10.5951/jresematheduc.47.1.0028

Estrella, S., Alvarado, H., Olfos, R., y Retamal, L. (2019). Desarrollo de la alfabetización probabilística: textos argumentativos de estudiantes. *Revista Paradigma, 40*(1), 280-304.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children.* Reidel. https://doi.org/10.1007/978-94-010-1858-6

Fischbein, E. (1987). *Intuition in science and mathematics: An educational approach.* Springer.

Fischbein, E., y Schnarch, D. (1997). The evolution with age of probabilistic intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*(1), 96-105. https://doi.org/10.2307/749665

Fulmer, G. W. (2014). Undergraduates' attitudes toward science and their epistemological beliefs: Positive effects of certainty and authority beliefs. *Journal of Science Education and Technology, 23*(1), 198-206. https://doi.org/10.1007/s10956-013-9463-7

Gardner, M. (1959). Mathematical games. *Scientific American, 219*, 180-182.

Garfield, J., y Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting research and teaching practice.* Springer.

Gómez, E., Batanero, C., y Contreras, J. M. (2014). Conocimiento matemático de futuros profesores para la enseñanza de la probabilidad desde el enfoque frecuencial. *Bolema: Boletim de Educação Matemática, 28*, 209-229. https://doi.org/10.1590/1980-4415v28n48a11

Gómez, E., Contreras, J. M., y Batanero, C. (2015). Significados de la probabilidad en libros de texto para educación primaria en Andalucía. En C. Fernández, M. Molina y N. Planas (Eds.), *Investigación en Educación Matemática XIX* (pp. 69-72). SEIEM.

Kahneman, D., Slovic, P., y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809477

Kahneman, D., y Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454. https://doi.org/10.1016/0010-0285(72)90016-3

Kahneman, D., y Tversky, A. (1982). Variants of uncertainty. *Cognition, 11,* 143-157. https://doi.org/10.1016/0010-0277(82)90023-3

Ministerio de Educación de Chile. (2015). Nuevas Bases Curriculares y Programas 7° básico a 2° año de Educación Media. Autor. http://www.curriculumenlineamineduc.cl/605/articles-34960_Bases.pdf

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics.* Autor.

Nisbett, R., y Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgments.* Prentice Hall.

Rodríguez, F., Díaz, D., y Vásquez, C. (2018). Evaluación de la alfabetización probabilística del profesorado en formación y en activo. *Estudios Pedagógicos, 46*(1), 135-156. https://doi.org/10.4067/S0718-07052018000100135

Sanabria, G., y Núñez, F. (2017). La probabilidad como elemento orientador de la toma de decisiones. *Revista Digital: Matemática, Educación e Internet, 17*(2), 1-13. https://doi.org/10.18845/rdmei.v17i2.3079

Serrano, L., Batanero, C., Ortiz, J. J., y Cañizares, M. J. (1998). Concepciones de los alumnos de secundaria sobre modelos probabilísticos en las secuencias de resultados aleatorios. *Suma, 36*, 23-32.

Sharma, S. (2014). Teaching probability: A socio-constructivist perspective. *Teaching Statistics, 37*(3), 78-84. https://doi.org/10.1111/test.12075

Tauber, L.; y Olesker, L. (2014). Significados dados a los fenómenos aleatorios en el contexto de la enseñanza media uruguaya. En: L. Tauber (Ed.), *Actas de V Jornadas de Educación Matemática y II Jornadas de investigación en Educación Matemática.* Facultad de Humanidades y Ciencias. Universidad Nacional del Litoral.

Tversky, A., y Kahneman, D. (1974). Judgement under uncertainity: Heuristics and biases. *Science, 185*, 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Vásquez, C., y Alsina, A. (2017). Lenguaje probabilístico: un camino para el desarrollo de la alfabetización probabilística. Un estudio de caso en el aula de Educación Primaria. *Bolema: Boletim de Educação Matemática, 31*, 454-478. https://doi.org/10.1590/1980-4415v31n57a22

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

https://doi.org/10.46219/rechiem.v13i4.79

CONOCIMIENTO SOBRE EL MUESTREO EN ESTUDIANTES CHILENOS AL TÉRMINO DE LA EDUCACIÓN ESCOLAR

CHILEAN STUDENTS' UNDERSTANDING OF SAMPLING AT THE END OF SCHOOL EDUCATION

Karen Ruiz-Reyes
kruiz@institutocomercialosorno.cl
Instituto Comercial de Osorno, Osorno, Chile

Felipe Ruz
felipe.ruz.a@pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Elena Molina-Portillo
elemo@ugr.es
Universidad de Granada, Granada, España

José M. Contreras
jmcontreras@ugr.es
Universidad de Granada, Granada, España

RESUMEN

En la actualidad, las nociones de inferencia estadística se incorporan dentro de las orientaciones curriculares de diversos países, incluido Chile, donde se le atribuye gran importancia a la noción de muestra y a su selección (muestreo). Por ello, este artículo tiene por objetivo indagar qué comprenden los estudiantes chilenos que cursan su último año de escolaridad sobre los métodos de muestreo. Se analizan las respuestas de 148 estudiantes chilenos de último año de secundaria (grado 12, 17-18 años) a un ítem de respuesta abierta sobre la identificación del sesgo de un método de muestreo. Los resultados muestran que los escolares prefieren el método de muestreo estratificado frente a otros y tienen dificultades para reconocer sesgos en la selección de muestras concretas.

PALABRAS CLAVE:
Técnicas de muestreo, Sesgos del muestreo, Educación secundaria.

ABSTRACT

At present, statistical inference has been incorporated into the curricular guidelines in several countries, including Chile, where great importance is attached to the notion of sample and its selection (sampling). This article aims to investigate what Chilean students understand about sampling methods at the end of their school education. The responses of 148 high school students (grade 12, 17-18 years old) to an open-ended item about bias identification in a sampling method are analyzed. The results show that high school students prefer the stratified sampling method and have difficulties in recognizing sampling bias in the selection of specific samples.

KEYWORDS:
Sampling Method, Sampling Bias, High School.

Recibido: 28 de Mayo de 2021, Aceptado: 27 de Julio de 2021

1. Introducción

Al analizar los documentos curriculares se puede apreciar que, durante los primeros niveles de escolaridad, la atención se centra en estudiar temas vinculados a la estadística descriptiva, mientras que los contenidos de inferencia estadística se han destinado para los últimos cursos de la trayectoria escolar, situación que puede evidenciarse en países como Estados Unidos, España y Chile (Common Core State Standards Initiative [CCSSI], 2010; Ministerio de Educación y Ciencia de España [MEC], 2015; Ministerio de Educación de Chile [MINEDUC], 2009, 2015a, 2015b). No obstante, en la literatura especializada sobre enseñanza de la estadística se ha resaltado el rol mucho más amplio y profundo de la inferencia para su tratamiento en la matemática escolar (Batanero, 2013; Makar y Ben-Zvi, 2011), sugiriendo su promoción desde los primeros años de educación obligatoria (Ben-Zvi et al., 2015; Meletiou-Mavrotheris y Paparistodemou, 2015). Entre las principales ideas de la inferencia estadística se distingue la noción de muestra y los métodos de muestreo, que se consideran conceptos elementales básicos para continuar con el estudio de temas más avanzados como los intervalos de confianza y contrastes de hipótesis. Por ello, Burrill y Biehler (2011) hacen hincapié en que los errores de comprensión sobre el muestreo pueden proyectarse en los contenidos posteriores.

Profundizando en la realidad chilena, las directrices curriculares (MINEDUC, 2009, 2015a, 2015b) que se resumen en la Tabla 1, representan que el concepto de *muestra* (entendida como un subconjunto de la población) se oficializa a partir del séptimo año de enseñanza básica (grado 7, 12-13 años), junto a otros conceptos como dato, población y variable estadística. Además, se promociona el estudio de la aleatoriedad de una muestra y se orienta a la reflexión intuitiva sobre su representatividad y la posibilidad de estimar resultados poblacionales a partir de ella.

Tabla 1. Objetivos de aprendizaje relacionados con el muestreo en el currículo chileno

Nota. Elaboración propia a partir de MINEDUC (2009, 2015a, 2015b).

Curso	**Objetivos de Aprendizaje (OA) / Aprendizaje Esperado (AE)**
Séptimo básico (12-13 años)	OA 15: Estimar el porcentaje de algunas características de una población desconocida por medio del muestreo. OA 16: Representar datos obtenidos en una muestra mediante tablas de frecuencias absolutas y relativas, utilizando gráficos apropiados, de manera manual y/o con software educativo.
Segundo medio (15-16 años)	OA 11: Utilizar permutaciones y la combinatoria sencilla para calcular probabilidades de eventos y resolver problemas.
Cuarto medio (17-18 años)	AE 14: Verificar mediante ejemplos concretos que la media de muestras aleatorias del tamaño n, extraídas de una población, se distribuye aproximadamente normal, si se aumenta el tamaño de la muestra.

En segundo medio (grado 10, 15-16 años) se profundiza en conceptos como el tamaño muestral, promoviendo el uso de distintas técnicas de conteo (permutaciones, variaciones y combinaciones) en el sorteo al azar, con o sin reposición. Mientras que en cuarto año medio (grado 12, 17-18 años) se promueve la realización de conjeturas sobre el tipo de distribución al que tienden las medias muestrales, con lo que se introduce la noción de distribución muestral y su relación con la distribución normal (MINEDUC, 2015b).

En este contexto, destacamos que en las directrices presentadas en la Tabla 1 no se explicita el trabajo con los distintos tipos de técnicas de muestreo, ni sus sesgos, por lo que surge la inquietud de indagar sobre qué comprenden los estudiantes respecto a estos conceptos. Desde esta perspectiva, en este trabajo se reportan los resultados de analizar las respuestas proporcionadas por una muestra de 148 estudiantes chilenos del último año de educación escolar obligatoria (grado 12, 17-18 años) a un ítem de respuesta abierta sobre la comprensión del muestreo y el uso de sesgos en su aplicación. Para ello, en lo que sigue, se sintetizan los principales antecedentes que sustentan este estudio.

2. Antecedentes

Dentro de la exhaustiva revisión bibliográfica realizada acerca del muestreo, podemos mencionar que la literatura científica existente, con estudios que involucren estudiantes en edad escolar, es muy escasa, incluso a nivel latinoamericano. Por ello hemos seleccionado algunas investigaciones que, a pesar de su antigüedad, abordan este tema y nos han servido de referencia para estructurar los antecedentes del estudio aquí expuestos.

Heitele (1975) distingue al muestreo como uno de los conceptos más relevantes entre las ideas estocásticas fundamentales. A pesar de ello, en la literatura se resalta que la investigación sobre el desarrollo de la comprensión del muestreo en escolares ha sido en general escasa (Makar y Rubin, 2018; Meletiou-Mavrotheris y Paparistodemou, 2015).

Por una parte, Jacobs (1997) analizó la comprensión informal sobre el muestreo en distintos grupos de alumnos estadounidenses de cuarto y quinto grado (9-11 años) al evaluar los métodos de muestreo en situaciones concretas y las conclusiones recogidas de ellas. Los niños utilizaron diversos criterios de evaluación según el contexto y el tipo de método de muestreo (restringido, autoseleccionado o aleatorio). Se destacan mayores dificultades para detectar sesgos con métodos de muestreo autoseleccionados que en los métodos de muestreo restringido, porque en las situaciones problemáticas presentadas a los estudiantes ellos consideraban que la autoselección era la manera más justa de seleccionar una muestra. Además, los estudiantes preferían el muestreo aleatorio estratificado frente al muestreo aleatorio simple.

Watson y Moritz (2000) desarrollaron una marco jerárquico o progresivo del razonamiento sobre el muestreo en escolares a partir de entrevistas a 62 estudiantes australianos de distintos grados. Años más tarde, Watson (2004) complementa estos resultados con un estudio longitudinal basado en entrevistas a 22 de los participantes de Watson y Moritz (2000) y otros 16 estudiantes de secundaria, después de tres o cuatro años de la consulta inicial. Los participantes prefieren la selección de muestras según algún procedimiento sesgado o intencional sobre las seleccionadas mediante un método de muestreo aleatorio. Además, tras clasificar las respuestas según el marco propuesto, la autora destaca que los participantes mejoran su razonamiento entre ambas entrevistas. Sin embargo, destaca que las bases curriculares australianas tienen una tarea pendiente con el muestreo, ya que en ellas se le otorga poco protagonismo a su enseñanza y es un aspecto elemental para promover el razonamiento inferencial.

Posteriormente, Watson y Kelly (2005) profundizan en los resultados previos analizando el razonamiento estadístico de 639 alumnos de los grados 3, 5, 7 y 9 (8-14 años), donde se incluían seis preguntas sobre muestreo basadas en las propuestas por Jacobs (1997). Estas interrogantes tenían por objetivo identificar sesgos en la selección de una muestra en diversas situaciones problema. Los alumnos consideran que los métodos de muestreo restringidos eran más fáciles de utilizar en comparación a los métodos de muestreo de tipo aleatorio, y más de la mitad fue capaz de detectar sesgos en el proceso de selección de muestras según cada contexto.

Por otro lado, Mavrotheris y Paparistodemou (2015) analizaron el razonamiento inferencial informal de 69 alumnos de 4° a 6° grado (9-11 años) sobre el conocimiento previo de conceptos tales como: muestra, su tamaño, el método de selección de muestras y el uso de sesgos en el muestreo. Algunos estudiantes parecían ser conscientes del sesgo en las muestras descritas, pero descartaron como apropiado el *muestreo aleatorio simple* porque estaban preocupados de que pudiera conducir a resultados extremos argumentando que “es aleatorio, por lo que cualquier cosa podría suceder”. Otros prefirieron el muestreo estratificado, no para aumentar la representatividad de la muestra a seleccionar, sino para aumentar la “imparcialidad” del proceso según argumentos basados en el género (“los niños y las niñas deben estar representados por igual”). También expresaron su preferencia por el muestreo autoseleccionado sobre el muestreo aleatorio simple, pero no pudieron detectar el sesgo involucrado en la autoselección debido a que, para ellos, la participación voluntaria fue “la más justa”.

A partir de los antecedentes previos, reforzamos nuestro interés por indagar en la comprensión sobre el muestreo en estudiantes chilenos, que nos permita diagnosticar y describir el panorama actual del conocimiento sobre la materia y proyectar algunas implicaciones para su enseñanza. Por ello, a continuación, se detalla la metodología empleada, se caracteriza a los participantes y se describen las categorías consideradas para el análisis de las respuestas entregadas a un ítem que evalúa distinguir sesgos en la selección de una muestra.

3. Metodología

Dada la naturaleza descriptiva del estudio de la que forma parte este trabajo, es que se utiliza una metodología mixta (Hernández et al., 2014) que comprende un análisis cualitativo de las respuestas proporcionadas por los estudiantes, de acuerdo a una serie de categorías que se describen en el apartado *categorías de estudio*, y también un análisis cuantitativo de la frecuencia en que se observan dichas categorías en las respuestas de los sujetos participantes en este estudio.

3.1 Participantes y contexto

La muestra está constituida por 148 estudiantes de cuarto año medio (grado 12, 17-18 años), entre ellos 89 mujeres (60%) y 59 hombres (40%), pertenecientes a tres centros educativos de la ciudad de Osorno, Chile. Los participantes cursaban el primer semestre del año 2018, de los cuales 48 (32%) asisten a centros de dependencia particular pagada (privados), 50 (34%) a centros particulares subvencionados y 50 (34%) a centros municipales (públicos).

Esta muestra es *no probabilística* y fue seleccionada por medio de un *muestreo casual* o incidental ya que corresponde a individuos procedentes de los centros educativos a los cuales los investigadores poseían *facilidad de acceso* (Bisquerra, 2004). En cuanto a su formación previa, no se puede afirmar que los estudiantes que forman parte de este estudio hayan recibido formalmente instrucción sobre el concepto de muestra y sus propiedades, ya que este criterio no fue pesquisado debido a que no formaba parte de los objetivos de nuestra investigación (Ruiz-Reyes, 2020).

3.2 Descripción del ítem

El instrumento utilizado corresponde a un cuestionario que se enmarca en una investigación doctoral sobre la comprensión del muestreo en estudiantes chilenos (Ruiz-Reyes, Ruz, Molina-Portillo y Contreras, 2019; Ruiz-Reyes, 2020), en el cual se incluye el ítem utilizado en este reporte, presente en la Figura 1.

En el ítem analizado en este manuscrito, se expone una situación problema que describe el resultado obtenido en la aplicación de encuestas realizadas a diferentes grupos de niños en un contexto escolar, en el que los métodos de muestreo y los resultados variaron, mientras que el tamaño de la muestra se mantuvo constante en 80 participantes. Se les pide a los alumnos que escriban sus respuestas completando la tabla que resume la información entregada en el enunciado y que además seleccionen cuál de los métodos expuestos es el método de muestreo que consideran más apropiado (Figura 1).

En este ítem se incluyen tres métodos de muestreo: aleatorio, restringido y autoseleccionado. En el método de *muestreo aleatorio* se asume que cada miembro de la población tiene la misma oportunidad de ser seleccionado. En el método de *muestreo restringido* se consideran grupos particulares de individuos que pueden ser más capaces de seleccionar una respuesta determinada y, en consecuencia, sesgar los resultados en una dirección particular. Mientras que en los métodos de *muestreo autoseleccionados* se permite que los participantes se seleccionen a ellos mismos y puedan afectar la representatividad de la muestra seleccionada (Jacobs, 1997). Caracterizando los métodos de muestreo expuestos en el enunciado del problema (Figura 1), se puede apreciar que el método de Andrea y Sara corresponde a un *muestreo restringido*; el método de Elena es un *muestreo autoseleccionado*; el método de Luis y el de Pedro son *muestreos aleatorios*; y más específicamente el método de Pedro corresponde a un muestreo estratificado.

Ítem. Los estudiantes de un colegio realizaron una encuesta para determinar el porcentaje de niños que reciclan en sus casas:

- Andrea preguntó a 80 estudiantes que son miembros del club de medioambiente.
- Elena envió un cuestionario a todos los niños del colegio y tomó los primeros 80 que contestaron.
- Pedro quiso el mismo número de niños y niñas. Así que preguntó a 5 niños y 5 niñas de cada curso para conseguir los 80 estudiantes que constituyen la muestra.
- Luis tenía los nombres de los 800 estudiantes en la escuela; puso cada nombre en un papel en un sombrero y sacó 80.
- Sara no conocía demasiados niños por lo que decidió encuestar a 80 niñas. Pero quería asegurarse de tener niñas de varias edades, por lo que tomó 10 niñas de cada nivel.

a. Completa la última columna donde preguntamos si el método de cada niño es apropiado.

Niño	**A quien pregunta**	**¿Te parece un método apropiado? ¿Por qué?**
Andrea	80 estudiantes del club de medioambiente.	
Elena	Envió un cuestionario a todos los niños del colegio y tomó los 80 que contestaron primero.	
Pedro	5 niños y 5 niñas de cada curso	
Luis	Puso el nombre de cada niño de la escuela (800) en un papel en un sombrero y sacó 80	
Sara	10 niñas de cada nivel.	

b. De acuerdo a los datos entregados en el enunciado, ¿Cuál crees que es la mejor manera de escoger a los niños para estimar el porcentaje de ellos que recicla en casa?

Figura 1. Ítem aplicado sobre el uso de sesgos en el muestreo

Nota. Ítem adaptado de Meletiou-Mavrotheris y Paparistodemou (2015) y de Watson y Kelly (2005).

Este ítem, previamente validado con estudiantes australianos y del Medio Oriente, ha sido adaptado del trabajo de Meletiou-Mavrotheris y Paparistodemou (2015) y de Watson y Kelly (2005), cuyas preguntas se han basado en las utilizadas por Jacobs (1997). Por ello destacamos la novedad de analizar su aplicación en Chile. Su objetivo es conocer la comprensión de los estudiantes sobre el uso de sesgos en el muestreo en situaciones problema de aplicación de encuestas. Los participantes respondieron las preguntas por escrito, como una actividad de la clase de Matemáticas, donde uno de los investigadores estuvo presente y apoyó la evaluación respondiendo consultas sobre la manera de completar el ítem y el cuestionario.

3.3 Categorías de análisis de las respuestas

Para cada uno de los métodos de muestreo señalados en el ítem aplicado (ver Figura 1), se establecen cuatro categorías de análisis:

- *Análisis apropiado*. Se describe la respuesta correcta a cada método de muestreo, es decir, los estudiantes identifican el sesgo del muestreo y emplean argumentos críticos apropiados.

- *Imparcialidad*. Se consideran las respuestas que evalúan cierta parte de la información, más específicamente en relación con la imparcialidad (que todos los sujetos de la muestra tengan la misma oportunidad de ser seleccionados) y el tamaño de la muestra.
- *Análisis inapropiado*. Se incluyen las respuestas que entregan una crítica inapropiada al método propuesto, las cuales se centran en la imprecisión percibida del método aleatorio, la injusticia o la imparcialidad, falta de oportunidad y el tamaño pequeño de la muestra.

• *Lógica inapropiada*. Considera los razonamientos erróneos que realizan los estudiantes.

A partir de estas categorías, en la Tabla 2 se proponen una serie de categorías creadas a partir de los resultados de Watson y Kelly (2005) y las respuestas obtenidas en este estudio, para analizar la información proporcionada por los estudiantes a la primera pregunta (a) del ítem de la Figura 1.

Tabla 2. Categorías de análisis pregunta a

Nota. Elaboración propia.

Método	**Análisis apropiado**	**Imparcialidad**	**Análisis inapropiado**	**Lógica inapropiada**
De Andrea	No apropiado porque se excluyen estudiantes	Tamaño no adecuado y método imparcial (grupo específico)	Apropiado porque es obvio que reciclan	Apropiado porque es un gran número de estudiantes
De Elena	No apropiado porque se excluyen estudiantes	Tamaño adecuado e imparcialidad inapropiada por quienes no contestan	Apropiado porque son quienes quisieron participar	Apropiado porque son los primeros en contestar
De Pedro	Apropiado por ser equitativo	Tamaño adecuado y selección justa	Inapropiado, debería consultar a todos los estudiantes	Inapropiado, se eligen muy pocos estudiantes
De Luis	Apropiado por ser un muestreo aleatorio	Tamaño adecuado	Inapropiado por ser muy aleatorio	Inapropiado por ser un método de selección muy complejo de ejecutar
De Sara	Inapropiado, deberían elegirse ambos sexos	Inapropiado por escoger solo mujeres	Apropiado por escoger sujetos con variedad de edades	Apropiado por coincidir en el interés de cuidar el medio ambiente

4. Resultados

En cuanto a los resultados obtenidos, en la Tabla 3 se muestran las frecuencias obtenidas de acuerdo a las categorías de análisis establecidas previamente en el apartado variables de estudio (Tabla 2). Se agrega en este caso una quinta categoría “No responde”, para cuantificar los casos sin respuesta.

Tabla 3. Número (porcentaje) de respuestas según categorías de estudio en pregunta a (Figura 1)

Nota. Elaboración propia.

Método	**Análisis apropiado**	**Imparcialidad**	**Análisis inapropiado**	**Lógica inapropiada**	**No responde**
De Andrea	38 (25,7)	61 (41,2)	22 (14,9)	19 (12,8)	8 (5,4)
De Elena	33 (22,3)	29 (19,6)	22 (14,9)	30 (20,3)	34 (23,0)
De Pedro	77 (52,0)	28 (18,9)	4 (2,7)	13 (8,8)	26 (17,6)
De Luis	71 (48,0)	14 (9,8)	8 (5,4)	39 (26,4)	16 (10,8)
De Sara	86 (58,1)	15 (10,1)	15 (10,1)	19 (12,8)	13 (8,8)

En cuanto a los resultados obtenidos en la categoría de "*análisis apropiado*", para el método de Andrea se aprecia que solo un 25,7% de los participantes identifica que es un *método de muestreo restringido* y que no es apropiado. En este caso, por ejemplo, se usan afirmaciones como: *"No me parece, porque debería haber preguntado a diferentes cursos"* (estudiante 20); "*no, Andrea debió preguntar a niños al azar"* (estudiante 55); *"no, no es equitativo"* (estudiante 100). Por otro lado, cuando se les presenta a los alumnos un método de *muestreo autoseleccionado*, como es el caso del método de Elena, solo un 22,3% responde realizando un *análisis apropiado*. En esta categoría, podemos encontrar argumentos como: *"No me parece el correcto, pues no está siendo justa con todos los estudiantes"* (estudiante 61); *"no es lo mejor porque así puede obtener respuestas de ciertos cursos no más, no te asegura que sean niños de todos los cursos"* (estudiante 77); *"no, ya que no es equitativo porque podrían haber contestado solo niños de un curso menor o mayor"* (estudiante 85).

En relación al método de Pedro (elige 5 niños y 5 niñas de cada curso), los alumnos consideran más adecuada su propuesta con un 52% de respuestas con análisis apropiados, usando argumentos basados en la equidad y justicia. Se hallaron en esta categoría las afirmaciones del tipo: *"Sí, porque fue equitativo"* (estudiante 11); *"sí, porque todos los cursos participaron"* (estudiante 48); *"sí, porque es justo"* (estudiante 145). Mientras, un 48% señala que es apropiado el *método de muestreo aleatorio* propuesto por Luis, justificando con expresiones como: *"Me parece apropiado porque de tal manera salen opiniones bastantes variadas"* (estudiante 32); *"sí, es una muestra aleatoria"* (estudiante 76).

Finalmente, cuando analizan el método de *muestreo restringido* correspondiente al método de Sara (elegir 10 niñas de cada curso), es el que mejor identifican ya que un 58,1% de los estudiantes indica que la muestra está sesgada, es decir, reconocen que no es correcto seleccionar una muestra donde los sujetos cumplan con una característica en particular y que, para realizar un método de muestreo apropiado, se deberían incluir tanto niños como niñas en dicha muestra. Algunos ejemplos de las justificaciones de los estudiantes fueron: *"No, ya que al haber elegido puras niñas puede que no puedan representar las ideas que tienen los varones"* (estudiante 16); *"no, porque los niños también deben dar su opinión"* (estudiante 24); *"no, debería ser parejo, preguntar a niños y niñas"* (estudiante 25); *"no, porque debió tomar a niños y niñas para hacerlo más completo"* (estudiante 92).

En cuanto a las respuestas clasificadas como *"imparcialidad"* en el método de Andrea, el 31,1% señala que no es un método apropiado porque *"al pertenecer al club de medio ambiente, todos los participantes contestan positivamente"* (estudiante 4). Solo un 10,1% señala que no es apropiado porque "*es un grupo específico"* (estudiante 13). A su vez, en el método de Elena, el 8,1% indica que es un método con tamaño de muestra apropiado porque *"se tomó la opinión a distintos alumnos"* (estudiante 18), y un 11,5% señala que no es apropiado porque *"la gente que piensa que era una mala idea no contestaría"* (estudiante 3). En cuanto al método de Pedro, un 11,5% de los estudiantes señala que es apropiado porque las *"respuestas serán variadas"* (estudiante 51) o un 7,4% indica porque *"se hace una imagen general"* (estudiante 21). En el caso de las respuestas al método de Luis, un 9,8% indica que es apropiado debido a que *"hay mucha gente"* (estudiante 11). Mientras que, en relación al método de Sara, un 10,1% sugiere que el método no es apropiado porque *"solo se escogieron niñas"* (estudiante 2).

En la categoría de *"análisis inapropiado"*, en el método de Andrea un 14,9% de los alumnos indica que es un método apropiado *"porque es obvio que reciclan"* (estudiante 1), es decir, en este caso no identifican que se está preguntando a un grupo sesgado. En el método de Elena, el mismo porcentaje de estudiantes menciona que el método es apropiado *"porque fueron los que quisieron participar"* (estudiante 5), en este caso, no logran ver que la representatividad de la muestra se ve comprometida por el método de selección. Para el método de Pedro, solo un 2,7% de los alumnos manifiesta que no es apropiado *"porque debería preguntarles a todos"* (estudiante 30), en este caso, no identifican el muestreo estratificado y quieren dar a todos los participantes la oportunidad de ser elegidos. En el método de Luis, un 5,4% de los alumnos considera que el método no es apropiado *"porque podría escoger a las personas equivocadas"* (estudiante 59), es decir, olvidan la importancia de la aleatoriedad en el proceso de selección de la muestra y tienen ideas como *"puede seleccionar a varios del mismo curso"* o *"puede elegir a varios que no reciclen"*. Finalmente, en el método de Sara, el 10,1% basa su opinión en que es un método apropiado debido a que *"selecciona niñas de distintas edades y obtendrá diferentes respuestas"* (estudiante 18), olvidándose que solo corresponde a un grupo en particular de sujetos y que su respuesta es sesgada.

Por lo que se refiere a la categoría *"lógica inapropiada"*, en el caso del método de Andrea un 12,8% de los estudiantes sugiere que es un método apropiado porque *"es una gran cantidad de estudiantes"* (estudiante 12). En el método de Elena, un 20,3% de los estudiantes menciona que la muestra ha sido bien seleccionada, ya que consideran que los participantes que respondieron primero y rápidamente son los que dan la mejor respuesta. A su vez, un 8,8% indica que el método de Pedro no es apropiado: *"Son muy pocos estudiantes"* (estudiante 1), y en el caso del método de Luis un 26,4% señala que el método no es apropiado porque *"es un método difícil"* o que *"lleva mucho tiempo"* (estudiante 10), no considerando el factor aleatorio de la elección de cada individuo que forma parte de la muestra. En el método de Sara, un 12,8% indica que es un método apropiado porque *"siempre están de acuerdo en cuidar el medio ambiente"* (estudiante 38). Y finalmente, se observa que un bajo porcentaje de estudiantes no responde (entre 5,4% y 10,8%), salvo con los métodos de Elena y Pedro, donde un 17,6% y un 23% de los participantes no entrega respuesta a esta situación (Tabla 3).

Por otro lado, en la Tabla 4 se presentan los resultados de la segunda pregunta (b) del ítem analizado (Figura 1), donde se les solicita a los participantes señalar cuál de los métodos propuestos es el más adecuado para escoger una muestra de estudiantes que recicla.

Tabla 4. Número (porcentaje) de participantes según tipo de respuestas a la pregunta b (Figura 1)

Nota. Elaboración propia.

Mejor método	**Andrea**	**Elena**	**Pedro**	**Luis**	**Sara**	**Otro**	**No responde**	**Total**
Frecuencia (%)	4 (2,7)	13 (8,8)	78 (52,7)	31 (20,9)	2 (1,4)	8 (5,4)	12 (8,1)	148 (100)

Podemos notar que un 73,6% de los participantes contesta señalando el método de Pedro o el método de Luis como el más apropiado (Tabla 4), prefiriendo con ello un *muestreo aleatorio* sobre los métodos sesgados. Si analizamos los resultados individuales del método elegido, se observa que 78 estudiantes (52,7%) prefiere el método estratificado de Pedro por sobre el método de muestreo aleatorio de Luis elegido por 31 estudiantes (20,9%).

Algunos estudiantes parecían ser conscientes del sesgo en el muestreo al analizar los métodos de Andrea, Elena o Sara, pero desestimaron como apropiado el muestreo aleatorio (método de Luis) porque estaban preocupados de que este pudiera conducir a resultados extremos: *"No me parece un método apropiado, porque si bien se seleccionan al azar no será posible controlar otras variables"* (estudiante 45), o mencionan argumentos relacionados a la imparcialidad: *"No, porque debería sacar una cantidad de alumnos y alumnas para quedar parejo"* (estudiante 132). También descartan el método aleatorio simple justificando que su uso es *injusto* debido a que algunos niños pudieron ser seleccionados y no querían participar, mientras que otros que no fueron seleccionados sí tendrían interés de hacerlo: *"No, porque al hacerlo por sorteo puede tener unas complejidades debido a que tal vez no todos estén interesados"* (estudiante 16).

5. Discusión

En este trabajo se han analizado las respuestas proporcionadas por una muestra de 148 estudiantes chilenos del último año de educación escolar obligatoria (grado 12, 17-18 años), a un ítem de respuesta abierta sobre la comprensión del muestreo y el uso de sesgos en su aplicación. En general, los resultados obtenidos respaldan los hallazgos descritos previamente por Jacobs (1997), Watson et al. (2003), Watson y Kelly (2005) y los de Meletiou-Mavrotheris y Paparistodemou (2015).

Respecto a la identificación de sesgos en distintos tipos de muestreo para la aplicación de una encuesta, se puede mencionar que un alto porcentaje de los alumnos (58,1%) identifica que un muestreo *restringido* (método de Sara) es un método sesgado debido a que solo se selecciona un grupo específico de personas (solo mujeres). En cambio, en el método de Andrea, que también corresponde a un muestreo restringido, ignoran que los alumnos seleccionados en la muestra corresponden a un grupo específico (pertenecen al club de medio ambiente) que no es representativo de la población, y solo un 25,7% señala que el método no es apropiado.

Otros alumnos prefirieron un *muestreo aleatorio*, como el método de Pedro (52%), no para aumentar la representatividad de la muestra, sino para aumentar la imparcialidad del proceso de selección, asegurando que los niños y las niñas estén representados por igual. Jacobs (1997) encontró que los niños seguían insistiendo en seleccionar un número igual de sujetos de cada género y concluyó que la preferencia por el muestreo estratificado puede estar vinculada a las concepciones estadísticamente no normativas de la imparcialidad de los niños. Este razonamiento se basa en creencias emotivas y personales de lo que constituye una encuesta justa (Watson et al., 2003).

En cambio, con el método de Luis (también de tipo aleatorio) solo el 48% de los estudiantes lo considera un método apropiado, pero un 26,4% da una respuesta con una lógica inapropiada, centrándose principalmente en cuestiones prácticas, como la dificultad de la ejecución del método de selección de la muestra y el tiempo que conllevará su aplicación.

A su vez, tuvieron dificultades para detectar el sesgo involucrado en la autoselección, como el método de Elena (22,3%), debido a la preocupación por cuestiones de *imparcialidad* de que tanto niñas y niños tuvieran la misma posibilidad de participar; o que la participación voluntaria es para ellos *más justa*, ya que a todos se les dio la misma oportunidad de responder el cuestionario. Resultados similares fueron reportados por Jacobs (1997) y Meletiou-Mavrotheris y Paparistodemou (2015).

En cuanto a los argumentos entregados para elegir alguno entre los métodos propuestos como el más apropiado, notamos que el papel del contexto juega un rol importante en el razonamiento del muestreo. Al respecto, Wroughton et al. (2013) sugieren que los alumnos que tienen opiniones firmes sobre un tema, evaluarán la validez de las conclusiones de un estudio basándose en si esa conclusión está o no de acuerdo con su opinión, en lugar de analizar la calidad del método de muestreo utilizado por medio de principios estadísticos. Esta situación también se observa en nuestro estudio, por ejemplo, al descartar al muestreo aleatorio simple como el más adecuado con argumentos relativos a la injusticia, como los reportados previamente por Watson et al. (2003), quienes señalan que a los estudiantes no les gusta generalizar a partir de una muestra aleatoria debido a la variabilidad inherente dentro de la población.

Para finalizar, podemos proyectar algunas líneas futuras de investigación en donde sería interesante efectuar un análisis comparativo de las respuestas al ítem considerando estudiantes de diversos grados, para identificar el nivel de escolaridad donde invertir esfuerzos iniciales en fomentar el conocimiento de los métodos de muestreo y los sesgos asociados. Además, podría analizarse el realizar un estudio longitudinal más amplio, considerando la elaboración y la implementación de una unidad didáctica, en que se aborden los temas de muestreo y sesgos, aplicando al inicio y al final de la instrucción el ítem expuesto en este manuscrito; como también considerar la puesta en común de las respuestas y los resultados con el grupo de estudiantes participantes. Además, sería interesante ampliar este estudio con algunas entrevistas, para profundizar en el razonamiento utilizado por los participantes al responder a cada una de las categorías.

Al ser este un estudio exploratorio descriptivo, los resultados obtenidos aportan valiosa información sobre la comprensión del muestreo de un grupo de estudiantes chilenos, como hemos señalado anteriormente, al existir pocos estudios que involucren estudiantes en edad escolar, consideramos que nuestros resultados son un significativo precedente para futuras investigaciones en este tema, ya que se ha podido establecer una caracterización inicial sobre este tema.

Referencias

Batanero, C. (2013). Del análisis de datos a la inferencia: Reflexiones sobre la formación del razonamiento estadístico. *Cuadernos de Investigación y Formación en Educación Matemática, 11*(8), 277-291.

Ben-Zvi, D., Bakker, A., y Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics, 88*(3), 291-303. https://doi.org/10.1007/s10649-015-9593-3

Bisquerra, R. (2004). *Metodología de la investigación educativa.* La Muralla

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). Springer. https://doi.org/10.1007/978-94-007-1131-0_10

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics.* National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics, 6*(2), 187-205. https://doi.org/10.1007/BF00302543

Hernández, R., Fernández, C., y Baptista, P. (2014). *Metodología de la investigación* (6.a edición). McGraw-Hill.

Jacobs, V. (1997, 24 al 28 de marzo). *Children's understanding of sampling in surveys* [Artículo de conferencia]. Annual Meeting of the American Educational Research Association, Chicago, IL, Estados Unidos.

Makar, K., y Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*, 1-4. https://doi.org/10.1080/10986065.2011.538291

Makar, K., y Rubin, A. (2018). Learning about statistical inference. En D. Ben-Zvi, K. Makar y J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 261-294). Springer. https://doi.org/10.1007/978-3-319-66195-7_8

Meletiou-Mavrotheris, M., y Paparistodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics, 88*(3), 385-404. https://doi.org/10.1007/s10649-014-9551-5

Ministerio de Educación de Chile. (2009). *Objetivos fundamentales y contenidos mínimos obligatorios de la enseñanza básica y media.* Unidad de Curriculum y Evaluación.

Ministerio de Educación de Chile. (2015a). *Bases curriculares: Matemática, educación media.* Unidad de Curriculum y Evaluación.

Ministerio de Educación de Chile. (2015b). *Programa de estudio para cuarto año medio. Matemática.* Unidad de Curriculum y Evaluación.

Ministerio de Educación y Ciencia de España. (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.* Autor.

Ruiz-Reyes, K., Ruz, F., Molina-Portillo, E. y Contreras, J. M. (2019). Comprensión del concepto de muestra por estudiantes chilenos de cuarto año medio de educación secundaria. En J.M. Contreras, M. Gea, M. López-Martín y E. Molina-Portillo (Eds), *Actas del Tercer congreso Internacional Virtual de Educación Estadística* (pp. 1-10).

Ruiz-Reyes, K. (2020). *Comprensión del muestreo por alumnos chilenos de educación secundaria* (Tesis doctoral). Universidad de Granada, Granada, España.

Watson, J. M. (2004). Developing reasoning about samples. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277-294). Kluwer Academic. https://doi.org/10.1007/1-4020-2278-6_12

Watson, J. M., y Kelly, B. (2005). Cognition and instruction: Reasoning about bias in sampling. *Mathematics Education Research Journal, 17*(1), 24-57. https://doi.org/10.1007/BF03217408

Watson, J. M., Kelly, B., Callingham, R., y Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. International *Journal of Mathematical Education in Science and Technology, 34*, 1-29. https://doi.org/10.1080/0020739021000018791

Watson, J. M., y Moritz, J. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*(1), 44-70. https://doi.org/10.2307/749819

Wroughton, J., McGowan, H., Weiss, L., y Cope, T. (2013). Exploring the role of context in students' understanding of sampling. *Statistics Education Research Journal, 12*(2), 32-58. https://doi.org/10.52041/serj.v12i2.303

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

ARTÍCULOS DE INVESTIGACIÓN

https://doi.org/10.46219/rechiem.v13i4.82

MEDIDAS DE TENDENCIA CENTRAL Y DISPERSIÓN MIRADAS DESDE UN DEPORTE TÍPICO CHILENO Y LA MODELACIÓN ESTADÍSTICA

MEASURES OF CENTRAL TENDENCY AND DISPERSION SEEN FROM A TYPICAL CHILEAN SPORT AND STATISTICAL MODELING

Elisabeth Ramos-Rodríguez
elisabeth.ramos@pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Natalia Alvarado-Garcés
natalia.alvarado.g@mail.pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Patrícia Vásquez
patricia.vasquez@pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Andrea Vergara
avergarag@ucm.cl
Universidad Católica del Maule, El Maule, Chile

RESUMEN

Este artículo presenta resultados de una propuesta de aula que considera la modelación estadística para la enseñanza de las medidas de tendencia central y dispersión, a partir del contexto que provee un deporte típico chileno. Desde un enfoque de investigación basado en el diseño, se elabora una propuesta de aula, implementada en estudiantes de 15 y 16 años. Su diseño promueve la modelación estadística, considerando como contexto un deporte nacional chileno, la rayuela, lo que permite abordar un problema cultural de interés para los estudiantes. Para el análisis de su implementación se articularon y adaptaron las fases de un ciclo de investigación estadística con las fases de un modelo asociado al proceso de modelación matemática. Los hallazgos de la implementación evidencian que los estudiantes lograron transitar exitosamente en las dos primeras fases del proceso de modelación, teniendo algunos de ellos más dificultad en las dos últimas fases. Este estudio permite evidenciar que aún es necesario continuar investigando en cómo la modelación de naturaleza estadística se inserta en el currículum matemático, y cómo estas pueden conjugarse, de modo de contar con criterios e indicadores claros para evaluar el desempeño de estudiantes en actividades que involucren análisis estadísticos en contexto.

PALABRAS CLAVE:
Modelación estadística, Medidas de tendencia central, Medidas de dispersión, Deportes típicos.

ABSTRACT

This article presents results of a classroom proposal that considers statistical modeling for the teaching of measures of central tendency and dispersion, based on the context provided by a typical Chilean sport. From a design-based research approach, a classroom proposal is developed, implemented in 15 and 16-year-old students. The design promotes statistical modeling, considering a Chilean national sport, hopscotch, as a context, which allows students to address a cultural problem of their interest. For the analysis of its implementation, the phases of a statistical research cycle were articulated and adapted to the phases of a model associated with the mathematical modeling process. The findings of the implementation show that students managed to successfully go through the first two phases of the modeling process, some of them having more difficulty in the last two phases. This study makes it possible to show that it is still necessary to continue investigating how modeling of a statistical nature is inserted in the mathematical curriculum, and how these can be combined, in order to have clear criteria and indicators to evaluate the performance of students in activities that involve statistical analysis in context.

KEYWORDS:
Statistical modeling, Measures of central tendency, Measures of dispersion, Typical sports.

Recibido: 30 de Mayo de 2021, Aceptado: 1 de Agosto de 2021

1. Introducción

En el ámbito escolar, la estadística y las probabilidades corresponden a un sector de la asignatura de Matemáticas donde las definiciones, propiedades y algoritmos se van presentando de manera progresiva. De acuerdo al Programa de Matemáticas de tercero y cuarto año (jóvenes entre 15 y 17 años de edad) de enseñanza media de Chile (Ministerio de Educación de Chile [MINEDUC], 2020), esta rama de la matemática promueve en los estudiantes maneras de pensar y de actuar en la toma de decisiones. Por otra parte, de acuerdo con la Asociación Americana de Estadística (ASA), la Estadística es "la ciencia de aprender a partir de datos y de medir, controlar y comunicar incertidumbre" (ASA, 2021, s. n., traducción propia), aunque en muchas ocasiones el término de estadística en el ámbito escolar se refiere a la información que proviene de tablas y gráficos (Saavedra, 2018).

El programa de estudios para tercer año medio en el sistema escolar chileno se refiere al uso de datos estadísticos y de modelos probabilísticos para la toma de decisiones, cuyo propósito es que los estudiantes interpreten datos en situaciones de incerteza. Para ello, el Ministerio de Educación de Chile estipula objetivos de aprendizaje, que consideran tanto el desarrollo de habilidades y destrezas, como el logro de conocimiento y comprensión: tomar decisiones en situaciones de incerteza que involucren el análisis de datos estadísticos con medidas de dispersión y probabilidades condicionales, tomar decisiones fundamentadas en evidencia estadística y/o en la evaluación de resultados obtenidos a partir de un modelo probabilístico y argumentar, utilizando lenguaje simbólico y diferentes representaciones, para justificar la veracidad o falsedad de una conjetura, y evaluar el alcance y los límites de los argumentos utilizados (MINEDUC, 2020).

Sin embargo, se observa que en Chile este contenido suele enseñarse al finalizar el año escolar, pasando a un segundo plano el análisis de los datos. Además, la enseñanza de las medidas de dispersión en los libros de texto se realiza prioritariamente presentando las definiciones y las propiedades y/o fórmulas (Del Pino y Estepa, 2019). Al respecto, Lee y Lee (2011) afirman que los profesores tienen dificultades en la comprensión de las medidas de tendencia central y de dispersión. Es importante resaltar que en Chile la formación de los docentes, en muchas de las instituciones formadoras, contempla dos asignaturas de Estadística y en el último tiempo se ha incorporado una relacionada con la didáctica del tema. Esta exigua dotación de asignaturas relacionadas con la estadística podría ser la razón de las dificultades que presentan algunos docentes para la enseñanza de la disciplina, dado que hay significativas diferencias entre el desarrollo del pensamiento estadístico y el matemático, pues el primero aspira a trabajar a partir de los datos según su contexto, mientras que el segundo se propone la abstracción de la realidad (Cobb y Moore, 1997). En este sentido, la estadística proporciona una metodología para abordar empíricamente información complicada e incierta, de una manera que sea útil y científicamente válida (Chambers, 1993).

Otro elemento interesante que revela el programa de estudio chileno está dado por la relación entre la habilidad de modelar y el aprendizaje de la estadística. El año 2012, en Chile, se puso en marcha el nuevo marco curricular para los cursos de primero a sexto año de educación básica (niños entre 6 y 11 años), el año 2016 se implementó desde séptimo básico a segundo año de educación media (alumnos entre 12 y 15 años) y el año recién pasado para los cursos de tercer y cuarto año de educación media (jóvenes entre 16 y 17 años). Estas bases curriculares (MINEDUC, 2012, 2016, 2020) buscan fomentar en los estudiantes el pensamiento matemático, para ello se hace necesario desarrollar las habilidades –relacionadas entre sí– de: resolver problemas, representar, modelar y argumentar y comunicar. De esta manera, en los últimos años se han presentado, en los diferentes niveles escolares, situaciones que conllevan la modelación, tomando en cuenta aquellas relacionadas en ámbitos sociales, reales y científicos, tal como se ha venido dando en el ámbito internacional (Solares et al., 2018).

En términos de estudios en educación matemática, la modelación ha ocupado un lugar preponderante desde hace ya dos décadas (véase, por ejemplo, Stillman et al., 2020). Aun así, Frejd y Bergsten (2018) mencionan que no hay un consenso entre los investigadores en educación matemática sobre lo que constituye a la modelación y esta se ha abordado de diferentes maneras en los documentos curriculares en el mundo, dándose un amplio espectro que va desde incorporar modelos matemáticos en otras asignaturas, además de Matemáticas, hasta no enseñar Matemáticas de manera independiente, sino que integrándola a otras materias, de tal manera de potenciar la modelación matemática. Esta problemática se hace latente también en la educación estadística, considerada en este estudio como una rama de la educación matemática, dada su inclusión en el currículum escolar en la asignatura de Matemática, por lo que el profesor que la enseña es el de Matemáticas.

La investigación en modelación matemática, en el ámbito de la educación matemática, ha sido prominente en la última década, pero aquella relacionada con la modelación estadística es limitada (Pfannkuch et al., 2018). Al respecto, Cobb (2007) sugiere que la modelación debería desempeñar un papel importante en la enseñanza de la estadística, afirmando que todos los métodos estadísticos se derivan de un modelo y que uno de los propósitos de la educación matemática es fomentar el pensamiento estadístico en los estudiantes a través de un razonamiento explícito entre el modelo y la realidad. El desarrollo de la modelación estadística ayuda a enfrentar las dificultades de aprendizaje de la estadística, permitiendo al estudiante desarrollar la capacidad de resolución de problemas (Tacoma et al., 2018).

Por su parte, Pfannkuch et al. (2018) argumentan la necesidad de desarrollar innovación en el campo de la educación estadística, en donde podemos destacar algunas propuestas como las de Biehler et al. (2018), Budgett y Pfannkuch (2018) y Kazak y Pratt (2017).

Atendiendo a la problemática expuesta, se presenta una propuesta de aula basada en la modelación estadística con la finalidad de que estudiantes le den significado a las medidas de tendencia central y de dispersión mediante su propia experimentación, poniendo en juego sus conocimientos previos, en especial en lo relativo al concepto de desviación media. La actividad matemática se desarrolla a partir de un deporte típico de Chile, la rayuela, que, de acuerdo con Cádiz (2018), tiene un encanto relevante en el país y que hoy por hoy ha ido perdiendo relevancia.

2. Marco de referencia

El marco de referencia de este estudio considera elementos para el diseño de propuestas de aula que involucra la modelación, los que orientarán la elaboración de la situación de aprendizaje sobre modelación estadística. Además, articula y adapta las fases del ciclo de investigación estadística de Wild y Pfannkuch (1999) con la propuesta de evaluación de las fases de un proceso de modelación matemática de Acebo-Gutiérrez y Rodríguez-Gallegos (2021). Esto aportará en la creación de categorías de análisis para la modelación bajo una conceptualización estadística. Se incorpora a este marco de referencia, una descripción de los elementos estadísticos que se pondrán en juego en la propuesta de aula. Del mismo modo, proveen herramientas para enfrentar el análisis de las producciones de los estudiantes.

2.1 Diseño de actividades de modelación

Para el diseño de actividades de modelación consideramos los principios relacionados con actividades de enseñanza para esta habilidad propuestos por Frejd y Bergsten (2018), así como lo planteado por Da Silva y Barbosa (2011), elementos que se describen a continuación.

En el estudio presentado por Frejd y Bergsten (2018) se presentan ocho principios (Tabla 1) para ser considerados en la planificación de actividades de enseñanza que involucren modelación, los que hemos considerado en el estudio.

Tabla 1. Descriptores de los ocho principios para la planificación de actividades escolares que involucran la modelación

Nota. Obtenido de Frejd y Bergsten (2018, pp. 125-126).

Principio	**Descriptores**
P1: Explicidad del objetivo de la actividad	Hacer explícito el objetivo del trabajo de modelación (describir, comprender, predecir, participar en una discusión crítica); cómo una descripción matemática/recreación del problema puede contribuir al objetivo.
P2: Centrarse en el problema	Definir el problema y considerar todo el problema durante el trabajo de modelación.
P3: Nivel del sistema	Ver el problema como enmarcado a través de un sistema, considerando que los sistemas parciales se modelan por separado.
P4: Supuestos y simplificaciones	Cuestionar por qué las variables elegidas son importantes.
P5: Relevancia y precisión de los datos	Considerar la calidad de los datos utilizados para estimar la influencia de las variables seleccionadas dentro del modelo.
P6: Papel de la tecnología	Constatar que el uso de tecnología es crucial para el trabajo de modelación.
P7: Cómo funciona el modelo	Revisar cómo funciona el modelo cuando se está probando (validación); si es posible probarlo o no, y qué tan útil y efectivo es para su propósito.
P8: Comunicación	Discusión o negociaciones entre los participantes a lo largo de la actividad de modelación.

Por otro lado, bajo una perspectiva sociocrítica, Da Silva y Barbosa (2011) proponen para el diseño de actividades de modelación tres aspectos: a) la presentación de una situación que aborde un problema socialmente relevante para los estudiantes, b) que los estudiantes tengan una participación activa en la construcción del modelo, y c) que el docente sea un mediador.

2.2 Ciclo de investigación en estadística y modelación estadística

En Wild y Pfannkuch (1999) encontramos un modelo del ciclo de investigación en estadística, conocido por las siglas PPDAC que resumen sus etapas: Problema, Plan, Datos, Análisis y Conclusión (Figura 1). De acuerdo con los autores, un ciclo PPDAC se ocupa de abstraer y resolver un problema estadístico basado en un problema real más grande, dando cuenta de la realidad que enfrentan los estadísticos cuando la conclusión de una investigación genera preguntas que conducen a otra investigación.

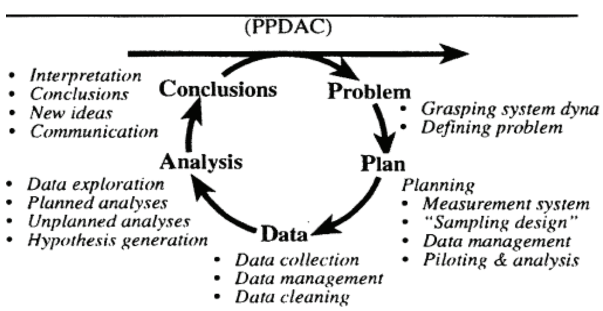


Figura 1. Ciclo de Investigación en Estadística
Nota. Obtenido de Wild y Pfannkuch (1999, p. 226).

Coincidimos con Wild y Pfannkuch (1999) en que se construyen y se usan modelos para comprender y predecir el comportamiento de la realidad y que, en relación a la estadística y la modelación estadística, al igual que la modelación matemática, simplifican la realidad, pero consideran con mayor énfasis la información que se tiene de ella. En la Figura 2 se encuentra el proceso de modelación estadística que dan estos autores y que hemos considerado en este estudio, el cual es un proceso de abstracción que supone las concepciones estadísticas del problema, donde influye la forma en que se recogen los datos sobre el sistema y se analizan.

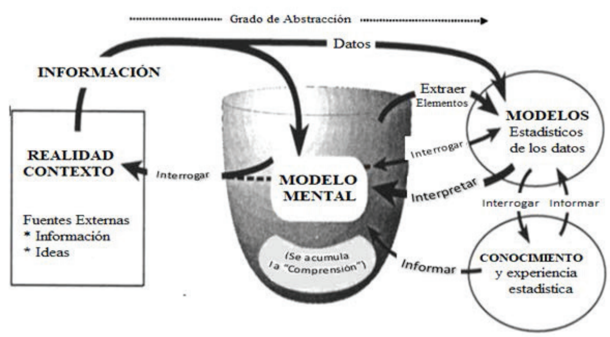


Figura 2. Proceso de Modelación Estadística
Nota. Traducción propia de Wild y Pfannkuch (1999, p. 230).

La Figura 2 ilustra la forma en que aprendemos sobre la realidad en contexto a medida que avanza una investigación estadística. La *comprensión* de la realidad del contexto se construye a partir de modelos mentales de esta. Estos modelos se basan en información de la realidad del contexto, por ejemplo, incorporando *conocimiento experto*. En un mundo ideal, estaríamos comprobando continuamente la idoneidad del mapeo entre modelo y realidad al *interrogar* la realidad del contexto. Parte de la información que buscamos y/o obtenemos de la realidad del contexto son datos estadísticos. Construimos modelos estadísticos para obtener información a partir de esta *interpretación* que retroalimenta el modelo mental.

2.3 Evaluación de la Modelación Matemática en fases

Acebo-Gutiérrez y Rodríguez-Gallegos (2021) proponen el diseño y validación de una rúbrica para la evaluación de modelación matemática en alumnos de secundaria, que nos brinda orientaciones precisas respecto de cómo distinguir las fases que tienen lugar cuando los estudiantes trabajan de forma individual para resolver un problema de modelación. Para efectos de este análisis se considerarán las descripciones generales de las fases, sin considerar los indicadores por nivel de logro, como se puede ver en la Tabla 2.

Tabla 2. Descripción de las fases para evaluar la resolución de problemas de modelación matemática
Nota. Adaptado de Acebo-Gutiérrez y Rodríguez-Gallegos (2021, p. 23).

Fases	**Descripción**
Formulación	Identifica el problema o situación del mundo real. Identifica las partes o datos relevantes del problema para su solución.
Resolución	Determina variables y parámetros para construir un modelo matemático. Genera un modelo matemático para representar el problema. Realiza cálculos y resuelve el modelo matemático.
Interpretación	Formula explicaciones. Hace supuestos y reconoce limitaciones.
Validación	Contrasta los resultados con la realidad. Reflexiona sobre otras formas de resolver el problema o desarrollar las soluciones existentes de diferentes maneras.

2.4 Fases en un ciclo de modelación estadística

Considerando la naturaleza específica del ciclo PPDAC, el proceso de modelación estadística propuesto por Wild y Pfannkuch (1999) y las recomendaciones para evaluar un proceso de modelación matemática, se proponen cuatro fases para la modelación estadística, de tal manera que estas puedan servir como categorías para clasificar las respuestas de los estudiantes cuando se enfrentan a una situación de modelación estadística. Esta propuesta reorganiza las fases del PPDAC según las cuatro fases de la propuesta de Acebo-Gutiérrez y Rodríguez-Gallegos (2021). Las fases *Problema y Dato* del ciclo PPDAC se enmarcan en la fase de *Formulación*, la fase *Planificación* del ciclo PPDAC se combina con la fase *Resolución*, la fase de *Análisis* del ciclo PPDAC complementa y amplía la fase *Interpretación*, finalmente la fase *Conclusión* del ciclo PPDAC se articula con la fase *Validación*. Esta reestructuración puede observarse en la Tabla 3.

Tabla 3. Descripción de las fases propuestas para analizar un proceso de modelación estadística
Nota. Elaboración propia.

Fases	**Origen**	**Descripción**
Comprensión	Problema y datos (PPDAC) + Formulación (rúbrica)	Se identifica o define el problema o situación del mundo real y las partes o datos relevantes para la solución de este. Se recolectan y gestionan los datos.
Construcción y resolución	Plan (PPDAC) + Resolución (rúbrica)	Se planifica el sistema de medición, determinando las variables de interés. Se propone un modelo estadístico para representar el problema y se pone a prueba (pilotaje). Se organizan los datos y se realizan los cálculos.
Interrogación e interpretación	Análisis (PPDAC) + Interpretación (rúbrica)	Se analizan los resultados obtenidos, formulando explicaciones. Se interrogan las posibles hipótesis o bien se hacen nuevos supuestos a partir de la exploración de los datos. También se reconoce limitaciones o redirecciones para continuar investigando.
Elaboración de información	Conclusión (PPDAC) + Validación (rúbrica)	Se contrastan los resultados con la realidad y se elaboran conclusiones. Se reflexiona sobre otras formas de resolver el problema. Se identifican y comunican las nuevas ideas construidas.

2.5 Conceptos basales: medidas de tendencia central y de dispersión

Las medidas de tendencia central son usadas para describir el centro o la localización de este en un conjunto de datos. Estas corresponden a la *moda*, la *mediana* o la *media aritmética*. La moda se calcula para datos numéricos, ordinales o nominales, la mediana para datos numéricos u ordinales y la media, para datos numéricos. De acuerdo con Saavedra (2018), la moda y la mediana no usan toda la información, ya que corresponden respectivamente al valor que más se repite y al valor que está justo al medio o al promedio de los dos datos centrales (cuando el número de datos es par). Por otra parte, una muestra podría no presentar una moda o bien ser bimodal, y la media es una medida sensible a los valores extremos, a diferencia de la mediana, que es más estable frente a cambios en los valores extremos o a la presencia de datos atípicos. Además, aunque la media considera todos los valores de la muestra, no tiene por qué ser uno de los valores de la muestra. Otra característica que tiene la media es que, si se tienen dos muestras distintas con la misma cantidad de datos, entonces la media de la muestra formada por ambas muestras es la suma de cada una de las medias, característica que no se cumple para la moda ni para la mediana.

Otro aspecto a considerar, de acuerdo con Saavedra (2018) en relación a la distribución de frecuencias, es que si esta distribución es:

- *Bastante simétrica*, entonces la media, la mediana y la moda son muy parecidas.
- *Asimétrica a la izquierda*, entonces media < mediana < moda.
- *Asimétrica a la derecha*, entonces moda <mediana < media.

Estas medidas son recomendadas para inferir el comportamiento de variables en poblaciones y muestras.

Las medidas de dispersión complementan a las medidas de tendencia central, siendo esenciales de considerar en la distribución de datos. A su vez, estas resultan ser útiles para comparar distribuciones y comprender los riesgos en la toma de decisiones. De acuerdo con De Veaux et al. (2003), *la dispersión o variabilidad* es la razón de ser de la estadística, donde la dispersión es la diferencia entre el valor observado y el valor verdadero del fenómeno en cuestión. Así, las medidas de dispersión "indican, hablando de un modo general, la distancia promedio de los valores muestrales al centro de la distribución" (Saavedra, 2018, p. 114).

3. Metodología

En este apartado se presentan los elementos metodológicos que guían el estudio y, posterior a ello, la propuesta de aula que se diseña, implementa y analiza.

3.1 Elementos metodológicos del estudio

Este estudio, de carácter cualitativo, busca recoger datos mediante la observación del comportamiento y expresiones de los sujetos puestos frente a un estímulo, buscando el sentido y significado que ellos le dan a la información obtenida, a través de los datos; en nuestro caso, la implementación de una propuesta de enseñanza. Se ciñe al enfoque de *investigación basada en el diseño*, dado que permite diseñar y evaluar intervenciones educativas, con el fin de resolver situaciones complejas, buscando generar y promover un conjunto de construcciones teóricas fundamentadas en contextos naturales. Esta investigación se centrará en explicar por qué el diseño (que considera la modelación estadística) funcionó y de qué manera puede ser adaptado a otras circunstancias.

Consideramos las etapas propuestas por Plomp (2010) para una investigación de diseño: teórica, experimental y terminal. La fase teórica está centrada en el desarrollo de un marco teórico o conceptual, en nuestro caso en relación a la modelación estadística y las medidas de tendencia central y de dispersión. En la fase experimental se diseña la propuesta de prototipo inicial, se realiza la intervención y una evaluación formativa de esta, de manera cíclica. Por último, en la fase terminal concluimos si la intervención cumple con las especificaciones predeterminadas.

Para la fase experimental se creó una situación de aula, en donde los estudiantes recogen los datos mediante su experimentación, dando sentido al proceso de modelación estadística. Esta se presenta en la subsección siguiente de este apartado.

La experimentación se realiza el año 2020 en situación de pandemia, por lo que la propuesta de aula se implementa de forma remota mediante un trabajo grupal previo y una clase virtual desde la plataforma *Zoom*, con una duración de 40 minutos. La implementación se llevó a cabo con un grupo de 28 estudiantes de tercer año medio (15 y 16 años) de un colegio particular subvencionado científico-humanista de la Región Metropolitana de Chile. Es importante destacar que los estudiantes, en el área de probabilidad y estadística, cuentan con conocimientos de gráficos de barras, diagramas, medidas de centralización, medidas de tendencia central para datos cualitativos y cuantitativos, medidas de tendencia central para datos agrupados y no agrupados y sus respectivas organizaciones en tablas de distribución de frecuencias y gráficos. Sin embargo, no han abordado los conceptos de dispersión de datos y de desviación estándar.

Se consideraron dos semanas de aplicación; en la primera de ellas se plantea la problemática a trabajar y los alumnos diseñan el juego (*la rayuela*) de forma asincrónica, a la semana siguiente se recopilan datos obtenidos por los estudiantes, para luego plantear las preguntas que provocan que emerja un modelo, su resolución y validación.

Los datos fueron recogidos a través de diversos canales. Uno de ellos corresponde a la *observación de las clases*, donde se obtiene *registro escrito* del chat por el cual los estudiantes se comunican durante la clase online, se obtiene un *registro visual*, ya sea en video o fotografías, de las construcciones y experimentaciones de los estudiantes, los cuales son enviados a la profesora. Por otra parte, está el *registro escrito* de un informe obtenido por los estudiantes una vez realizada la experimentación y vaciado de la información, los cuales se extraen dando respuesta a las preguntas planteadas, las que tienen como fin obtener información a partir de sus propios análisis y conclusiones.

El método utilizado para el análisis es el método de *análisis de contenido*, donde las unidades de análisis han sido las intervenciones de los alumnos en la clase y las respuestas a la tarea propuesta en el informe grupal. Las categorías de análisis corresponden a las fases de la *Modelación Estadística* propuestas en el marco teórico (Tabla 3). Cada fase tiene sus respectivos indicadores, a su vez los indicadores tienen cuatro niveles para su medición, dependiendo del nivel de presencia de dicha categoría en la producción de los estudiantes, donde 1 es el valor mínimo y 4 es el valor máximo de presencia (Tabla 4).

Tabla 4. Indicadores y niveles para las categorías de análisis utilizadas en el estudio.

Nota. Elaboración propia.

Categorías	**Comprensión**		**Construcción y resolución**			**Interrogación e interpretación**		**Información**	
Indicadores	Identifica el problema o situación del mundo real y construye el modelo mental	Recolecta datos de la realidad	Determina variables y parámetros para construir un modelo estadístico	Genera un modelo estadístico para representar el problema	Realiza cálculos y resuelve el modelo estadístico	Formula explicaciones	Contrasta los resultados con la realidad	Reflexiona sobre otras formas de resolver el problema o desarrollar las soluciones	Presenta los resultados del estudio
Niveles	1/2/3/4	1/2/3/4	1/2/3/4	1/2/3/4	1/2/3/4	1/2/3/4	1/2/3/4	1/2/3/4	1/2/3/4

3.2 El diseño del prototipo de intervención

El objetivo principal de la tarea es la indagación y la recogida de datos por parte de los estudiantes y la utilización de las medidas de tendencia central y de dispersión a partir de un deporte nacional: la rayuela.
Según consta en la Biblioteca Nacional de Chile (Ministerio del Deporte de Chile, 2014), la ley número 20.777, promulgada el 17 de septiembre de 2014, señala: "Declárase la actividad deportiva de la rayuela como deporte nacional, que se disputa por puntos, consistente en el lanzamiento de tejos, desde distancias prefijadas, hacia una superficie determinada, atravesada por una lienza a alcanzar" (p.1).
En Báez et al. (2020) se indica que en tiempos de la Colonia era un juego de la calle, donde niños y adultos se entretenían y se señala las normas de este, las que se dan en la Figura 3.

Se parte con los tejos; existen tejos cilíndricos y redondos metálicos, según el lugar donde se juegue dependerá la forma del tejo; de Arica a Chillán se usa el tejo cilíndrico y en el Sur, la forma redonda conocido como paica.

Los jugadores deben elegir la forma de los tejos antes de empezar el juego, no pueden jugar con tejos diferentes, debe ser el mismo tipo para todos, solo puede variar el peso de estos, lo que queda a criterio del jugador, pero siempre manteniendo la distancia de la rampa de lanzamiento a la lienza.

La rampa de lanzamiento es donde los jugadores tiran los tejos y tiene una distancia de 14 metros hasta el receptáculo, acá es donde caen los tejos; es un marco construido de barro con una leve inclinación para que los jugadores puedan ver mejor la lienza, esta lienza divide el marco y es en ella donde los tejos deben caer para que se produzca la quemada, lo que le otorga los puntos al jugador. La meta del juego son doce puntos, el primer jugador o equipo que llegue a este número es el que gana.

Figura 3. Instrucciones del deporte nacional chileno la rayuela

Nota. Obtenido de Báez et al. (2020, p. 371)

Considerando lo anterior, se elabora la propuesta de aula, de tal manera que pueda ser aplicada y realizada en conjunto con los estudiantes en contexto online. De esta forma, se presentó a los estudiantes la construcción del cajón de rayuela y, mediante la completación de las tablas de distribución de frecuencias, ellos podían concluir algunas situaciones, entregando fotografías y un informe realizado por ellos en el que se ponía de relieve los descubrimientos obtenidos. A continuación, en la Figura 4, se presenta la tarea diseñada en relación a la rayuela.

La Rayuela

Instrucciones

i) En cada partido, deben jugar dos personas, quienes tienen que ubicarse a dos metros de la caja, tal como lo indica la Figura.

ii) Cada jugador debe hacer 10 lanzamientos, en cada uno de ellos debe medir la distancia entre la ubicación donde quedó el tejo (la marca que hay en la caja) al punto medio del hilo (el punto que está ubicado al centro del hilo de la caja) y anotar en la tabla dicha longitud en cm.

Número de lanzamiento	Jugador 1 Distancia entre la marca al punto medio del hilo en cm	Jugador 2 Distancia entre la marca al punto medio del hilo en cm
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Desafío

a) Juega a la rayuela con integrantes de tu familia, recuerda que debes registrar las medidas en la tabla.

b) En grupo de cuatro personas deben presentar sus tablas y a partir de ellas, responder las siguientes preguntas

1. ¿Cuál creen que fue el mejor jugador? ¿Por qué?
2. ¿Cuál debería ser el jugador ganador? ¿Por qué?
3. ¿Cómo determinaron al ganador de la partida? Explique el procedimiento con sus propias palabras
4. Si hubiera una expresión matemática que representara al ganador ¿Cuál creen que podría ser dicha expresión, para determinar el ganador en las futuras partidas de rayuela?

Figura 4. Tarea escolar presentada a los estudiantes

Nota. Elaboración propia.

La tarea escolar planteada pretende que emerja de los estudiantes un concepto estadístico a través de la intuición y la interpretación de los datos recogidos, considerando la participación activa de estos en la construcción del modelo y una actuación del docente como mediador.

Como ya se mencionó, en la propuesta de aula se han considerado los principios de diseño para una actividad de modelación planteados por Frejd y Bergsten (2018). En la tarea escolar se hace explícito el objetivo del trabajo de modelación, se define el problema y es considerado durante el trabajo de modelación. Con las preguntas planteadas en el enunciado de la tarea, se busca que los estudiantes cuestionen por qué las variables elegidas son importantes, indaguen sobre la calidad de los datos utilizados para estimar la influencia de las variables seleccionadas dentro del modelo, y sean capaces de delimitar cómo funciona el modelo cuando se está probando; si es posible probarlo o no, y qué tan útil y efectivo es para su propósito.

Desde la mirada sociocrítica bajo las directrices de la modelación propuesta por Da Silva y Barbosa (2011), la innovación considera la presentación de una situación que aborda un problema socialmente relevante para los estudiantes, la rayuela, como deporte nacional típico, donde se vela por que los estudiantes tengan una participación activa en la construcción del modelo, y donde el docente toma un rol de mediador.

4. Resultados

Los resultados se presentan en relación a la implementación de la propuesta de aula que considera la modelación estadística para la enseñanza de las medidas de tendencia central y dispersión, a partir del contexto que provee un deporte típico chileno, aplicada a estudiantes de 15 y 16 años. La modalidad de trabajo de los estudiantes fue en grupos de cuatro, donde, para efectos del presente escrito, se les nombró como grupo A, B, C, D, E, F y G.
Al tabular las respuestas, según las fases de la Modelación Estadística planteadas, se obtiene el resumen que se presenta en la Tabla 5.

Tabla 5. Tabulación de respuestas según fases del proceso de modelación

Nota. Elaboración propia.

Grupo	**Comprensión**		**Construcción y resolución**			**Interrogación e interpretación**		**Información**	
	Identifica el problema o situación del mundo real y construye el modelo mental	Recolecta datos de la realidad	Determina variables y parámetros para construir un modelo estadístico	Genera un modelo estadístico para representar el problema	Realiza cálculos y resuelve el modelo estadístico	Formula explicaciones	Contrasta los resultados con la realidad	Reflexiona sobre otras formas de resolver el problema o desarrollar las soluciones	Presenta los resultados del estudio
A	4	4	4	4	4	4	2	2	2
B	4	4	4	4	4	4	4	4	4
C	4	4	4	4	2	1	1	1	1
D	4	4	4	4	4	2	1	1	1
E	4	4	4	4	4	3	1	1	1
F	4	4	4	4	4	4	4	4	4
G	3	3	3	4	2	1	1	1	1

Es posible observar que seis de los siete grupos logran transitar en su completitud por la fase de comprensión (fase 1) y la fase de construcción y resolución (fase 2); en consecuencia, un solo grupo transitó parcialmente por ambas fases. Por otra parte, en el caso de la fase de interpretación (fase 3), esta es lograda en su totalidad por dos de los siete grupos, los cuales son los que finalmente completan a su vez la última fase (de información, fase 4), mientras que los otros grupos lograron un escaso o superficial cumplimiento de ellas. A continuación, se detallan los hallazgos para cada fase del proceso de modelación.

4.2.1 Fase 1: Comprensión

La primera parte del desafío de la tarea: J*uega a la rayuela con algún integrante de tu familia, completa la tabla señalada en las instrucciones y responde las preguntas siguientes: 1) ¿Cómo determinarían al mejor jugador? ¿Por qué?, 2) ¿Cómo determinarían al jugador ganador? ¿Por qué?*, nos instan a evidenciar si los estudiantes son capaces de reconocer las instrucciones entregadas por el docente, si comprenden el problema y si son capaces de recolectar los datos.

De la Tabla 1 se puede observar que todos los alumnos fueron capaces de atravesar satisfactoriamente esta fase. Llama la atención que la respuesta de dos de los siete grupos se relaciona con la distancia al hilo, como se ilustra en la Figura 5, donde se puede evidenciar que el grupo C fue capaz de identificar el problema, posicionándose sobre quién es el ganador y el mejor jugador, argumentando su decisión desde un razonamiento basado en la distancia al hilo en cada tirada y la cantidad de veces que acertaba el jugador en el hilo, sin emplear herramientas estadísticas.

1. ¿Cuál creen que fue el mejor jugador? ¿Por qué?
Respuesta: *Yo creo que Marcela, porque fue la que menos distancia tuvo del hilo y la que más veces dio en el centro.*

2. ¿Cuál debería ser el jugador ganador? ¿Por qué?
Respuesta: *Está entre Moira y Marcela, ya que, si sumamos los puntajes, Moira es la que menos obtuvo, pero Marcela es la que acertó más veces.*

Figura 5. Respuesta del grupo C sobre la comprensión del problema
Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

Otras respuestas que también aludieron a la distancia al hilo argumentaban desde otra perspectiva: cantidad de tiros acertados, como se puede ver en la Figura 6, donde el grupo G justifica considerando la cantidad de tiros acertados, considerando que el grupo ganador corresponde a aquel que obtiene mas veces el resultado de mayor puntuación, no tomando en cuenta la cantidad de veces que este no acierta, el cual en algunos casos puede ser mayor. Nuevamente la comprensión del problema no evidencia relación con elementos de estadística.

1. ¿Cuál creen que fue el mejor jugador? ¿Por qué?
Respuesta: *Yo creo que el mejor jugador fue el jugador 1, porque fue el que estuvo más cerca a la cuerda.*

2. ¿Cuál debería ser el jugador ganador? ¿Por qué?
Respuesta: *Debería ser el jugador 1, porque dio en el centro 2 veces.*

Figura 6. Respuesta del grupo G sobre la comprensión del problema
Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

Desde otra arista, es importante destacar aquellas respuestas dadas por los grupos que van más allá de la distancia al hilo, donde incorporan, en su comprensión del problema, justificaciones basadas en medidas de *tendencia central*, como es *el promedio de las distancias.* Destacamos la respuesta del grupo A (Figura 7), que alude al *"promedio de lejanías más bajo a los demás tiros respecto al oponente"*.

1. ¿Cuál creen que fue el mejor jugador? ¿Por qué?
Respuesta: *El mejor jugador fue el número 1, ya que tuvo más aciertos al hilo.*

2. ¿Cuál debería ser el jugador ganador? ¿Por qué?
Respuesta: *El jugador número 1 debería ganar, porque además de haber tenido más aciertos al hilo, tuvo un promedio de lejanía más bajo en los demás tiros respecto a su oponente.*

Figura 7. Respuesta del grupo A sobre la comprensión del problema
Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

Por último, resaltamos la respuesta del grupo E (Figura 8), quien de manera intuitiva hace emerger la noción de dispersión, a partir de expresiones como las de *"constantemente más cerca"* y *"más cerca"*, realizando además una comparación general.

1. ¿Cuál creen que fue el mejor jugador? ¿Por qué?
Respuesta: *El jugador número dos, ya que a diferencia de los otros dos jugadores (uno y tres), él estuvo constantemente más cerca del hilo durante todo el juego.*

2. ¿Cuál debería ser el jugador ganador? ¿Por qué?
Respuesta: *El jugador dos, porque durante todos sus turnos estuvo más cerca al hilo, incluso en el lanzamiento número siete logró lanzar el tejo justo al hilo.*

Figura 8. Respuesta del grupo E sobre la comprensión del problema
Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

De esta forma se visualiza la fase de comprensión como una instancia que permite a los estudiantes abordar el problema a partir de distintas posturas o fundamentos, las que les permiten avanzar a la fase dos, de construcción y resolución.

4.2.2 Fase 2: Construcción y resolución

La pregunta 3 de la tarea: *¿Cómo determinan al ganador de la partida?* es clave para evidenciar si los alumnos fueron capaces de transitar por la fase de resolución. De la Tabla 5 se puede visualizar que, si bien no todos los grupos lograron avanzar en plenitud, dos de ellos tuvieron algunas dificultades para realizar cálculos de forma correcta. Cinco de los siete grupos fueron capaces de determinar variables y parámetros para representar el problema, realizar cálculos y resolver el problema a partir de la generación de un modelo estadístico.

Naturalmente, el tipo de razonamiento que ocupan en esta fase depende de la forma de comprender el problema. Se observa que tres de los siete grupos (C, G y D) ocupan modelos básicos, como es el caso del grupo de C, quienes se centran en el máximo de las sumas de los puntajes de cada jugador (Figura 9).

> 3. ¿Cómo determinan al ganador de la partida? Expliquen el procedimiento con sus propias palabras.
> Respuesta: *Se suman todos los puntajes y el que menor valor obtuvo es el ganador.*

Figura 9. Respuesta del grupo C que evidencia su procedimiento para resolver. Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

Es relevante la respuesta del grupo E, puesto que utilizan el rango y el promedio de las distancias haciendo uso implícitamente de la desviación estándar, siendo estos modelos estadísticos que le permiten resolver el problema, como se ilustra en la Figura 10.

> 3. ¿Cómo determinan al ganador de la partida? Expliquen el procedimiento con sus propias palabras.
> Respuesta: *Comparamos los rangos entre las distancias en los lanzamientos de los jugadores y nos apoyamos al comparar los promedios de las distancias registradas, el jugador que haya registrado distancias menores con el hilo será el ganador.*

Figura 10. Respuesta que evidencia grupo E, del uso de más de una herramienta estadística. Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

Una respuesta interesante la presentan los integrantes del grupo F (Figura 11), quienes utilizan diversas estrategias y modelos estadísticos. Ellos inician su resolución comparando los puntajes de cada lanzamiento, luego consideran una medida de tendencia central, la media aritmética de las distancias. Esta media aritmética corresponde a un conocimiento previo de los estudiantes, el que es presentado como el "promedio de las distancias". Hay que hacer notar que, si bien este grupo se acerca a la noción de dispersión, el mayor promedio no da cuenta de una mayor dispersión de los datos. Para evidenciar una medida de dispersión tendrían que haber calculado, por ejemplo, la desviación media, sacando un promedio de las diferencias positivas entre cada distancia asociada a un lanzamiento y el promedio, situación que fue aclarada por la docente en la institucionalización.

> Respuesta: *Primero analizamos los resultados y los comparamos respecto al número de lanzamiento, luego sacamos los promedios del primer jugador y del segundo, nuevamente comparamos los resultados y el promedio del jugador 1, con valor de 9,23, es menor al promedio del jugador 2 con valor de 16,69, significando que la media aritmética del jugador 1 es menor a la del jugador 2. Esto nos lleva a afirmar que la dispersión de datos es mayor en el jugador 2 que en el jugador 1.*

Figura 11. Respuesta del grupo F que evidencia el uso de más de una herramienta estadística. Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

La resolución que cada grupo realiza les permite avanzar con la fase de interrogación e interpretación, lo que se detalla a continuación.

4.2.3 Fase 3: Interrogación e interpretación

La fase 3 de la Modelación Estadística se evidencia a partir de la segunda parte de la pregunta *4: Expliquen el procedimiento con sus propias palabras.* De los siete grupos, tres de ellos pudieron abordar esta fase de manera satisfactoria, principalmente en lo que respecta a la formulación de sus argumentos.

Al indagar en la respuesta del grupo G (Figura 6), se puede inferir que ellos fueron capaces de explicar su procedimiento para resolver el problema a partir de la comparación de la media aritmética obtenida en cada caso y argumentar que es necesario ocupar otro estadígrafo para concluir, observando las limitaciones de la primera de ellas (la media aritmética).

En la Figura 12 está el razonamiento del grupo B, quienes justifican su elección a partir de la media aritmética y la relación con el contexto y las jugadas acertadas.

> *Fácilmente podemos decir que el ganador absoluto fue el jugador 2, ya que fue el que más cerca llegó al hilo según el promedio anteriormente calculado, y además que, en uno de los lanzamientos, pudo llegar con un tejo al hilo (tiro 9).*

Figura 12. Respuesta del grupo B en relación a la fase de interpretación. Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

4.2.4 Fase 4: Información

Esta fase es posible evidenciarla a partir de las respuestas formuladas en la pregunta 4: *Si hubiera una expresión matemática que permita representar al ganador, ¿cuál creen que podría ser dicha expresión,* *para determinar el ganador en las futuras partidas de rayuela?*

La respuesta del grupo B se muestra en la Figura 13. En ella se puede observar que ellos fueron capaces de contrastar su respuesta con la realidad, considerando la relación entre la media aritmética y el lanzamiento del tejo. Además, reflexionan sobre otras formas de abordar el problema, recurriendo a otras herramientas matemáticas, pensando en supuestos como los gráficos o funciones, pero afirman que no lo saben con certeza, reconociendo sus limitaciones en términos de conocimientos para generar una argumentación más robusta y precisa.

4) Si hubiera una expresión matemática que representara al ganador, ¿cuál creen que podría ser dicha expresión, para determinar el ganador en las futuras partidas de rayuela?
Respuesta: *La verdad es que investigamos un poco, buscamos juegos matemáticos para tomar ideas, pero no se nos ocurre cuál puede ser esa expresión. Probablemente se debería establecer con el uso de gráficos y funciones matemáticas, tema que hemos estado estudiando en clases pero que aún no dominamos.*

Figura 13. Respuesta del grupo B

Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

Por otra parte, destacamos la respuesta del grupo F (Figura 14) quienes fueron capaces de contrastar sus resultados con la realidad y asociar otro tipo de soluciones desde otras áreas del saber, como la física.

4) Si hubiera una expresión matemática que representara al ganador, ¿cuál creen que podría ser dicha expresión, para determinar el ganador en las futuras partidas de rayuela?
Respuesta: *Creemos que una especie de cálculo que tenga que ver con frecuencias y rango, por ejemplo, calcular la diferencia del mayor y menor resultado que obtuvo un jugador, así se comparan los rangos de los participantes y, dependiendo del que cuenta menor rango, tiene más posibilidades de ganar.*

Por ejemplo:
Rango 1 = 56-0 *Rango 1 < Rango 2*
Rango 2 = 67-0 *56 < 67*

Pensamos que otra forma de expresión matemática, que sería en cierta medida más extensa, tendría que ver con las mediciones del centro de gravedad del deportista, ángulo de tiro, fuerza de impulso y velocidad de lanzamiento del tejo, dado que estos conceptos influyen en el juego. Por lo tanto, su implementación en el cálculo no estaría errónea.

Es posible que en este tipo de juego exista una técnica para obtener buenos lanzamientos, la costumbre y dedicación también son factores que influyen en ello.

Figura 14. Respuesta del grupo F

Nota. Elaboración propia a partir de las respuestas obtenidas de los estudiantes.

5. Conclusiones

Este trabajo muestra los resultados del diseño e implementación de una propuesta de aula basada en la modelación estadística con la finalidad de que estudiantes le den significado a las medidas de tendencia central y de dispersión. El diseño se desarrolló considerando los ocho principios para la planificación de actividades escolares que involucran la modelación (Frejd y Bergsten, 2018), decantando por un contexto que incluye un deporte nacional chileno, la rayuela, que aborda un problema socialmente relevante para los estudiantes, donde ellos tienen una participación activa en la construcción del modelo, y el docente es un mediador (Da Silva y Barbosa, 2011).

Los resultados de la implementación de la propuesta se analizaron a partir de la adaptación y articulación entre el ciclo de investigación y el proceso de modelación estadística (Wild y Pfannkuch, 1999) con la propuesta de evaluación para el ciclo de modelación matemática (Acebo-Gutiérrez y Rodríguez-Gallegos, 2021). Los hallazgos evidencian que los estudiantes lograron transitar exitosamente en las dos primeras fases del proceso de modelación, teniendo algunos de ellos dificultades en las dos últimas fases. A pesar de esto, fueron capaces de reconocer los conceptos involucrados, y mediante la interpretación y el análisis de datos avanzaron hacia la toma de decisiones.

Si bien la actividad de modelación fue pensada para que los estudiantes interactuaran físicamente, ello no fue posible por el contexto sanitario de Chile del año 2020, solo daba la posibilidad de que los estudiantes se relacionaran de manera virtual. Esto pudo influir en que no todos los grupos pudiesen llegar a las últimas fases de la modelación. Aun así, es importante destacar que la actividad está pensada con la intencionalidad de contar con una instancia de interacción familiar, con un contexto que se conozca (medianamente) dando relevancia a un deporte nacional chileno que ha perdido fuerza entre las nuevas generaciones.

Esta propuesta de innovación tuvo como propósito proponer una situación socialmente relevante a los estudiantes, junto con el de generar y fomentar procesos de modelación con ellos, que les permitan acercarse a la estadística a través de otra arista y desde la realidad, realidad que les provoque interactuar con su entorno cercano y reconocer las matemáticas y/o la estadística en ello.

Esta innovación pretende mostrar un escenario en donde se conjuguen elementos del ámbito tradicional chileno (la rayuela) con la enseñanza de la estadística, articulación que promueve la modelación estadística, de manera que pueda ser un insumo para otros docentes en la generación de conocimiento y habilidades de otras generaciones de estudiantes.

El estudio evidencia que aún es necesario continuar investigando cómo la modelación, de naturaleza estadística, se inserta en el currículum matemático, de manera que se rescaten elementos como lo es el deporte nacional en la construcción de situaciones y conceptos matemáticos. Además, se deja entrever la necesidad de generar ambientes más cercanos al estudiante, teniendo en cuenta que, en la medida que los estudiantes avanzan hacia los últimos años de escolaridad, dejan de "jugar" para aprender, lo que puede provocar asignaturas rígidas o abstractas.

Por último, sostenemos la importancia de contar con actividades de modelación estadística que puedan conjugarse con las propuestas ministeriales, de modo de tener criterios e indicadores claros para evaluar el desempeño de los estudiantes en actividades que involucren análisis estadísticos en contexto.

Una de las implicancias prácticas de este estudio está en la posibilidad de aplicar la actividad de modelación propuesta en otros niveles de enseñanza, usando y adaptando las categorías de análisis como rúbrica para la evaluación del desempeño grupal de los estudiantes. Del mismo modo, el estudio ofrece un recurso concreto para instalar el trabajo con otras disciplinas, como historia y educación física, que se puede concretar en un ABP (aprendizaje basado en proyectos).

Algunas proyecciones del estudio pueden apuntar a que la propuesta de innovación incorpore otras preguntas que ayuden a los estudiantes a completar las fases en el ciclo de modelación. Por ejemplo, se podría incluir una situación hipotética, en la que dos de los jugadores lograran el mismo promedio de las distancias con el propósito de animar a los estudiantes a elaborar y comunicar nuevos argumentos para dirimir quién debe ser el ganador. Esto permitiría continuar investigando sobre cómo surge la necesidad de utilizar medidas de dispersión en este tipo de situaciones de modelación estadística y cómo estas se acoplan o coordinan con las medidas de tendencia central.

Agradecimientos

Este trabajo ha sido parcialmente financiando por la beca para Magister Profesionales de la Educación 50210054 de la Agencia Nacional de Investigación y Desarrollo, ANID, Chile.

Referencias

Acebo-Gutiérrez, C. J., y Rodríguez-Gallegos, R. (2021). Diseño y validación de rúbrica para la evaluación de modelación matemática en alumnos de secundaria. *Revista Científica, 40*(1), 13-29. https://doi.org/10.14483/23448350.16068

ASA. (2021). *American Statistical Association*. https://www.amstat.org/

Báez, P., Berríos, F., Rubio, M., Muñoz, M. J., y Guevara, J. (2020). *Tras las memorias de la comunidad de Petorca* [Trabajo Final de grado, Universidad Católica de Valparaíso]. Casiopea. https://wiki.ead.pucv.cl/Proyecto_de_titulo_-_Tras_las_memorias_de_la_comunidad_de_Petorca

Biehler, R., Frischemeier, D., Reading, C., y Shaughnessy, J. (2018). Reasoning about data. En D. Ben-Zvi, K. Makar y J. Garfield, *International handbook of research in statistics education* (pp. 139-192). Springer International Handbooks of Education. https://doi.org/10.1007/978-3-319-66195-7

Budgett, S., y Pfannkuch, M. (2018). Modelling and linking the Poisson and exponential distributions. *ZDM Mathematics Educatioin, 50*, 1281-1294. https://doi.org/10.1007/s11858-018-0989-2

Cádiz, O. (2018). *Juegos tradicionales y populares en Chile*. Ediciones Universitarias de Valparaíso. Pontificia Universidad Católica de Valparaíso. Chile.

Chambers, J. M. (1993). Greater or lesser statistics: A choice for future research. *Statistics and Computing, 3*(4), 182-184. https://doi.org/10.1007/BF00141776

Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology innovations in statistics education, 1*(1), 1-15. https://doi.org/10.5070/T511000028

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

Da Silva, J., y Barbosa, J. (2011). Modelagem Matemática: as discussões técnicas e as experiências prévias de um grupo de alunos. *Boletim de Educação Matemática, 24*(38), 197-218.

Del Pino, J., y Estepa, A. (2019). Análisis de la enseñanza de las medidas de dispersión en los libros de texto. Avances de I*nvestigación en Educación Matemática, 16*, 86-102. https://doi.org/10.35763/aiem.v0i16.232

De Veaux, R. D., Bock, D. E., y Velleman, P. (2003). *Intro Stats*. Addison-Wesley.

Frejd, P., y Bergsten, C. (2018). Professional modellers' conceptions of the notion of mathematical modelling: ideas for education. *ZDM, 50*(1), 117-127. https://doi.org/10.1007/s11858-018-0928-2

Kazak, S., y Pratt, D. (2017). Pre-service mathematics teachers' use of probability models in making informal inferences about a chance game. *Statistics Education Research Journal, 16*(2), 1-18. https://doi.org/10.52041/serj.v16i2.193

Lee, H. S., y Lee, J. T. (2011). Enhancing prospective teachers' coordination of center and spread: A window into teacher education material development. *The Mathematics Educator, 21*(1), 33-47.

Ministerio de Educación de Chile. (2012). *Programas de estudio, matemáticas*. Autor.

Ministerio de Educación de Chile. (2016). *Programas de estudio, matemática*. Autor.

Ministerio de Educación de Chile. (2020). *Programas de estudio, tercero medio para matemáticas.* Autor.

Ministerio del Deporte de Chile (2014). *Ley 20777, reconoce a la rayuela como deporte nacional.* https://www.bcn.cl/leychile/navegar?idNorma=1067843

Pfannkuch, M., Ben-Zvi, D., y Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM, 50*(7), 1113-1123. https://doi.org/10.1007/s11858-018-0989-2

Plomp, T. (2010). Educational Design Research: An introduction. En T. Plomp y N. Nieveen (Eds.), *An introduction to Educational Design Research* (pp. 11-50). Netherland Institute for Curriculum.

Saavedra, E. (2018). *Contenidos Básicos de Estadística y Probablidad.* Editorial USACH.

Solares, A., Preciado, A. P., Peña, F., Ortiz, A., Sandoval, M., Soriano, R., Carrión, V., y Fuentes, M. (2018). Tendencias en Modelación Matemática en Latinoamérica. En T. E. Hodges, G. J. Roy y A. M. Tyminski (Eds.), *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 88-100). University of South Carolina & Clemson University.

Stillman G. A., Kaiser, G., y Lampen, E. (2020). Sense-Making in Mathematical Modelling and Applications Educational Research and Practice. En G. Stillman, G. Kaiser y C. Lampen (Eds.), *Mathematical Modelling Education and Sense-making. International Perspectives on the Teaching and Learning of Mathematical Modelling* (pp. 15-29). Springer. https://doi.org/10.1007/978-3-030-37673-4_2

Tacoma, S., Sosnovsky, S., Boon, P., Jeuring, J., y Drijvers, P. (2018). The interplay between inspectable student models and didactics of statistics. *Digital Experiences in Mathematics Education, 4*(2), 139-162. https://doi.org/10.1007/s40751-018-0040-9

Wild, C., y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

https://doi.org/10.46219/rechiem.v13i4.84

SIGNIFICADO DE LA MEDIA, MEDIANA Y MODA EN TEXTOS ESCOLARES DE SÉPTIMO BÁSICO

MEANING OF THE MEAN, MEDIAN AND MODE IN SEVENTH GRADE SCHOOL TEXTS

Jaime I. García-García
jaime.garcia@ulagos.cl
Universidad de Los Lagos, Osorno, Chile

Ingrid B. Urrutia Leiva
urrutiaingrid79@gmail.com
Universidad de Los Lagos, Osorno, Chile
Colegio Bicentenario Crucero "Oscar Daniel", Río Bueno, Chile

Sebastián H. Vásquez Chicao
sebastianhvasquez@gmail.com
Universidad de Los Lagos, Osorno, Chile
Liceo San Sebastián, Osorno, Chile

Elizabeth Hernández Arredondo
elizabeth.hernandez@ulagos.cl
Universidad de Los Lagos, Osorno, Chile

RESUMEN

En este trabajo se presenta un estudio del significado de la media, mediana y moda en una muestra de dos textos escolares de séptimo año de Educación Básica de Chile. Para el desarrollo del estudio, hemos considerado algunas herramientas de análisis del Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos, identificando la presencia de cada elemento del significado (situaciones-problema, procedimientos, conceptos-definición, propiedades, elementos lingüísticos y argumentos) de la media, mediana y moda en los textos escolares. Los resultados evidencian la ausencia de diversos elementos en la presentación de la media, mediana y moda en los textos escolares chilenos de séptimo básico; en consecuencia, se deducen algunas sugerencias de los elementos de estas medidas de tendencia central que consideramos necesarios para su incorporación en las aulas para lograr su enseñanza más significativa.

PALABRAS CLAVE:
Enfoque ontosemiótico, Medidas de tendencia central, Educación Básica, Libros de texto.

ABSTRACT

This paper presents a study of the meaning of the mean, median and mode in a sample of two seventh-year textbooks of Basic Education in Chile. For the development of the study, we have considered some analysis tools of the Onto-semiotic Approach to Mathematical Cognition and Instruction, identifying the presence of each element of meaning (problem-situations, procedures, concepts-definitions, properties, linguistic elements and arguments) of the mean, median and mode in the school texts. The results show the absence of various elements in the presentation of the mean, median and mode in Chilean school textbooks for the seventh grade; consequently, some suggestions are deduced from the elements of these measures of central tendency that we consider necessary for their incorporation in the classrooms to achieve their most meaningful teaching.

KEYWORDS:
Onto-semiotic Approach, central tendency measures, primary education, textbooks.

Recibido: 4 de Junio de 2021, Aceptado: 14 de Julio de 2021

1. Introducción

Hace más de 25 años, la Estadística se ha incluido paulatinamente en el currículo de matemáticas de diversos países (Vásquez y Alsina, 2014) con el fin de desarrollar en el estudiante competencias para la recolección y organización de datos; la construcción, lectura e interpretación crítica de tablas y gráficos estadísticos; el cálculo e interpretación de medidas estadísticas; el análisis de la variabilidad; la predicción y toma de decisiones con base en los datos; entre otras. En general, para desarrollar la alfabetización estadística en el estudiante.

La alfabetización estadística, en términos de Gal (2004), se refiere a la capacidad de las personas para interpretar y evaluar críticamente información estadística, argumentos relacionados con los datos o fenómenos estocásticos, que pueden encontrar en diversos contextos; y la capacidad para discutir o comunicar sus opiniones respecto a dicha información, cuando sea relevante.

Del Pino y Estrella (2012) señalan que un ciudadano alfabetizado estadísticamente debería ser capaz de leer e interpretar representaciones estadísticas (tablas y gráficos) y medidas de resumen que se presentan en los medios de comunicación. En ese sentido, las medidas de tendencia central (MTC), parte de las medidas de resumen, son consideradas como elementos de la alfabetización estadística que todo estudiante debería aprender y comprender durante su educación.

Con respecto a los estudios sobre las MTC (media, mediana y moda), podemos señalar dos líneas de investigación: la primera enfocada en los errores y dificultades en la comprensión del significado de las MTC por parte de estudiantes y profesores de diversos niveles educativos (p. ej., Batanero, 2000; Cobo, 2003; Del Puerto et al., 2007; Escobedo y Mayén, 2018; Estrella, 2016; Mayén, 2009; Mayén et al., 2007); y la segunda centrada en el análisis de las MTC en libros de texto (p. ej., Cabrera, 2014; Carvalho y Gitirana, 2014; Cobo, 2003; Cobo y Batanero, 2004; Díaz-Levicoy et al., 2020; Estrella, 2008; Mayén, 2009; Ocoró y Ocoró, 2016).

Las investigaciones en torno al análisis del libro de texto radican en el reconocimiento de la importancia que tiene este recurso pedagógico en el proceso de enseñanza-aprendizaje de las matemáticas en las aulas escolares producto de una transposición didáctica (Chevallard, 1991), debido a que es la herramienta didáctica más usada por el profesor e influye en el aprendizaje de los estudiantes, así como su relación directa con la implementación del currículo (Herbel, 2007) y su contribución al éxito del cumplimiento y puesta en práctica de las directrices curriculares (Cantoral et al., 2015).

Por lo anterior, este estudio tiene como objetivo analizar el significado de la media, mediana y moda en una muestra de dos textos escolares chilenos de séptimo año de Educación Básica.

2. Referentes teóricos

2.1 El enfoque ontosemiótico (EOS)

Para el desarrollo de este estudio hemos utilizado algunas herramientas de análisis del modelo teórico denominado Enfoque Ontosemiótico (EOS) del Conocimiento y la Instrucción Matemáticos, propuesto por Godino y sus colaboradores (Godino et al., 2007, 2020).

El EOS es un modelo teórico que articula diversas perspectivas de las disciplinas y dimensiones relacionadas con el aprendizaje y enseñanza de las matemáticas para poder abordar problemas de tipo epistemológico, ontológico, semiótico-cognitivo y educativo (Pino-Fan, 2013). En el EOS se reconoce una doble naturaleza para las matemáticas: como sistema de objetos y como sistema de prácticas. La noción de práctica matemática se entiende como "toda actuación o expresión (verbal, gráfica, etc.) realizada por alguien para resolver problemas matemáticos, comunicar a otros la solución obtenida, validarla o generalizarla a otros contextos y problemas" (Godino y Batanero, 1994, p. 334); y la de objeto matemático como cualquier entidad material o inmaterial que interviene en la práctica matemática, apoyando y regulando su realización (Font et al., 2013).

La realización de una práctica matemática pone en funcionamiento conocimiento matemático. Este conocimiento se expresa en un lenguaje, que se puede descomponer en lenguaje verbal y lenguaje simbólico o gráfico. Los enunciados lingüísticos derivados de las prácticas matemáticas son, a su vez, la expresión de conceptos, proposiciones y procedimientos, y con base en estos, se constituyen argumentos que dan cuenta de la validez de las acciones.

De acuerdo con Font et al. (2007), de esta práctica matemática emergen seis tipos de objetos matemáticos primarios, denominados "elementos del significado", que corresponden a:

a) Situaciones-problema: referentes a aplicaciones extra matemáticas, tareas, ejercicios, cuestionamientos, entre otros, asociados al objeto matemático, en nuestro caso la media, mediana y moda.
b) Procedimientos: referentes a algoritmos, operaciones, técnicas de cálculo, entre otros, que se realizan en los diversos tipos de prácticas.
c) Conceptos-definición: introducidos mediante definiciones o descripciones, explícitas o implícitas, del objeto matemático (media, mediana, ...).
d) Propiedades/proposiciones: enunciados (declaraciones) sobre conceptos (postulados, teoremas, etc.).

e) Elementos lingüísticos: corresponden a términos, expresiones, notaciones, tablas, gráficos, etc., en sus diversos registros (oral, escrito, ...).
f) Argumentos: corresponden a enunciados (discursos) usados para probar o explicar las proposiciones y procedimientos, deductivos o de otro tipo; demostraciones que utilizamos para validar propiedades.

Estos elementos de significado están relacionados entre sí, formando configuraciones de redes de objetos intervinientes y emergentes de los sistemas de prácticas (Godino et al., 2007). De acuerdo con los autores, un sujeto comprende un determinado objeto matemático cuando lo usa de manera idónea en diferentes prácticas. Por ello, para el análisis de las MTC en los textos escolares de séptimo básico se considerará la tipología de objetos matemáticos primarios para identificar la presencia de cada elemento del significado de la media, mediana y moda.

2.2 Las medidas de tendencia central en las Bases Curriculares de Chile

El Ministerio de Educación de Chile (MINEDUC), en las Bases Curriculares de Educación Básica (MINEDUC, 2016, 2018) establece Objetivos de Aprendizaje (OA) que precisan la expectativa formativa que se espera que logren los estudiantes en cada asignatura, por cada año escolar. A partir de una revisión en estas bases, identificamos los OA que declaran el estudio de las MTC (Tabla 1).

Tabla 1. Objetivos de aprendizaje (OA) relacionados con las MTC en Educación Básica.
Nota. Elaboración propia a partir de MINEDUC (2016, 2018).

Grado escolar	**Rango etario**	**Objetivo de aprendizaje**
Quinto	10-11 años	23. Calcular el promedio de datos e interpretarlo en su contexto. (MINEDUC, 2018, p. 249)
Séptimo	12-13 años	17. Mostrar que comprenden las medidas de tendencia central y el rango: • Determinando las medidas de tendencia central para realizar inferencias sobre la población. • Determinando la medida de tendencia central adecuada para responder un problema planteado. • Utilizándolos para comparar dos poblaciones. • Determinando el efecto de un dato que es muy diferente a los otros. (MINEDUC, 2016, p. 109)

Como podemos observar, el estudio de las tres MTC se declara en séptimo básico; por ello, enfocamos el análisis en los textos escolares de dicho grado escolar.

3. Algunas investigaciones sobre análisis de las MTC en libros de textos

A continuación, se presentan algunas investigaciones centradas en el análisis de los elementos del significado de la media, mediana y moda en libros de texto, con el propósito de situar nuestro estudio.

En Cobo (2003) se presenta un estudio del significado institucional de las MTC en una muestra de 22 libros de texto de tercer (14-15 años) y cuarto (15-16 años) grado de Educación Secundaria Obligatoria (ESO) de España. Se analiza el tipo de problemas, algoritmos de cálculo, definiciones, propiedades, representaciones y argumentos, identificando aquellos elementos de significado más comunes presentados en la muestra de libros sobre las MTC, por ejemplo, obtener un elemento representativo de un conjunto de valores dados cuya distribución es aproximadamente simétrica, o bien, obtener un valor representativo de una colección de datos en situaciones en las que lo que interesa fundamentalmente es el valor dominante del conjunto. Con respecto al trabajo de las MTC, la autora señala que este le da mucha más importancia a la memorización de las definiciones y al cálculo de la media, mediana y moda, que al estudio de sus propiedades. Por su parte, Mayén (2009) analiza los elementos del significado de las MTC en dos libros de texto y un cuaderno de prácticas de matemáticas de tercer grado (14-15 años) de Educación Secundaria de México. Sus resultados concuerdan con los reportados por Cobo (2003), por ejemplo, en relación con los procedimientos más frecuentes (cálculo de la media, la mediana y la moda de una variable discreta con datos aislados), así como la escasez de aquellos de mayor complejidad tanto para la media como para la mediana.

Carvalho y Gitirana (2014) analizan los significados, propiedades y representaciones de la media aritmética en libros de texto de Brasil. Sus resultados evidencian que la mayoría de las actividades no promueven la reflexión de las propiedades; el significado con mayor presencia es la media como elemento representativo de un conjunto de datos, y la representación más utilizada es la lengua materna (lenguaje), considerada en el enunciado escrito, muchas veces ayudada con fórmulas y símbolos inherentes a la dimensión matemática de la media aritmética. En otra investigación, Ocoró y Ocoró (2016) analizan la forma en que se presentan las MTC en dos libros de texto colombianos de séptimo grado. Sus resultados evidencian que no se presentan todas las propiedades de cada MTC, no se trabaja la media ponderada, carecen de ejemplos e ilustraciones asociadas a sus actividades, y no ofrecen ejercicios de mayor dificultad que permitan al estudiante interiorizarse en estas medidas.

En el contexto chileno, Estrella (2008) analiza la transposición didáctica de las MTC en un libro de texto de séptimo básico, es decir, las transformaciones que se producen en los conocimientos sobre las MTC desde un texto de nivel universitario para ser enseñados en dicho grado. Sus resultados muestran que el libro de texto presenta imprecisiones con respecto a los términos y sus significados, las MTC son presentadas a través de fórmulas y las tareas se enfocan únicamente a su uso.

Con relación a lo anterior, se observa que son escasas las investigaciones enfocadas en el análisis de la media, mediana y moda, de manera conjunta, en los libros de textos. Bajo esta perspectiva, nuestro estudio aporta resultados de los elementos del significado de las MTC que se presentan en los textos escolares y se trabajan en las aulas chilenas de séptimo básico.

4. Metodología

Nuestro estudio se enmarca en una metodología cualitativa (Pérez-Serrano, 1994), de tipo descriptiva (Hernández et al., 2010), y se usa como método de investigación el análisis de contenido (López-Noguero, 2002), debido a que el objetivo es analizar cada elemento del significado (situaciones-problema, procedimientos, conceptos-definición, propiedades, elementos lingüísticos y argumentos) de la media, mediana y moda en los textos escolares de séptimo año de Educación Básica. La muestra seleccionada, no probabilística, es intencional y está formada por el texto del estudiante (TE) y el cuaderno de actividades (CA) de séptimo básico; ambos textos escolares siguen el marco curricular actual (MINEDUC, 2016) y son distribuidos gratuitamente para todos los centros educativos particulares y públicos con carácter obligatorio por parte del Ministerio de Educación de Chile (MINEDUC). En este sentido, estos textos escolares fueron elegidos por su amplio uso en las aulas chilenas.

Tabla 2. Textos escolares utilizados en el análisis.

Nota. Elaboración propia.

Texto escolar	Referencia
TE	Iturra, F., Manosalva, C., Romero, D., y Ramírez, M. (2020). *Matemática. 7° Básico. Texto del estudiante*. SM S.A.
CA	Arce, D. (2020). *Matemática. 7° Básico. Cuaderno de actividades*. SM S.A.

Para el análisis de los textos escolares seguimos el método propuesto por Cobo (2003): 1) lectura rigurosa de la Lección 17, Medidas de Tendencia Central, de cada texto escolar, con el propósito de comparar el contenido con cada elemento del significado de la media, mediana y moda, y con ello determinar su presencia (en caso de encontrar alguno nuevo se describe); 2) selección de imágenes para ejemplificar cada elemento del significado encontrado; 3) realización de un resumen escrito en el que se analizan los elementos identificados; 4) elaboración de tablas que resuman los resultados, y a su vez, obtención de conclusiones sobre los elementos del significado de las MTC en los textos escolares utilizados por los estudiantes de séptimo básico. Cabe señalar que el primer paso siguió un proceso sistemático y cíclico, cuya fiabilidad se aseguró mediante la comparación de los resultados del análisis realizado por cada uno de los autores, y en caso de desacuerdo, se realizó un nuevo proceso de análisis hasta llegar a un consenso. Para el análisis de contenido se consideró cada elemento del significado (situaciones-problema, procedimientos, conceptos-definición, propiedades, elementos lingüísticos y argumentos) de las MTC, a partir de la descripción de cada uno de ellos propuestos por Cobo (2003) y Mayén (2009), los cuales se describen y ejemplifican a detalle en el apartado de resultados.

a) Situaciones-problema (SP). a) Media (M): SPM1 [situación-problema asociada a la media, uno]. Estimar una medida a partir de diversas mediciones realizadas, en presencia de errores. SPM2. Obtener una cantidad equitativa al hacer un reparto para conseguir una distribución uniforme. SPM3. Obtener un elemento representativo de un conjunto de valores dados cuya distribución es aproximadamente simétrica. SPM4. Conocer el valor que se obtendrá con mayor probabilidad al tomar un elemento al azar de una población para una variable con distribución aproximadamente simétrica. SPM5. Comparación de dos distribuciones de datos con variables numéricas. b) Mediana (ME): SPME1. Encontrar un resumen estadístico de posición central, en situaciones en las que la media no es suficientemente representativa. SPME2. Encontrar un resumen estadístico de posición central para variables ordinales. SPME3. Efectuar comparaciones de dos o más colecciones de datos usando gráficos de caja. c) Moda (MO): SPMO1. Obtener como valor representativo de una colección de datos en situaciones en las que lo que interesa fundamentalmente es el valor dominante del conjunto. SPMO2. Encontrar el valor representativo en datos cualitativos.

b) Procedimientos (P). a) Media (M): PM1 [procedimiento asociado a la media, uno]. Cálculo de la media de una variable discreta con datos aislados. PM2. Cálculo de la media de una variable discreta con datos presentados en tablas de frecuencias. PM3. Cálculo de la media de una variable continua o discreta con datos agrupados en intervalos de clases. PM4. Cálculo gráfico. PM5. Cálculo con calculadora u ordenador. PM6. Inversión del algoritmo de cálculo de la media. PM7. Construir una distribución de media dada. b) Mediana (ME): PME1-PME2. Cálculo de la mediana con datos no agrupados en clases, número de datos impar y par, respectivamente. PME3-PME4.

Cálculo de la mediana a partir datos presentados en una tabla de frecuencias; casos de un número par e impar de valores, respectivamente. PME5. Cálculo de la mediana a partir de datos agrupados en clases. PME6. Cálculo de la mediana a partir de datos presentados en un gráfico. c) Moda (MO): PMO1. Cálculo de la moda en una variable discreta con datos aislados. PMO2. Cálculo de la moda en una variable discreta con datos presentados en una tabla de frecuencias. PMO3. Cálculo de la moda de una variable discreta o continua con datos agrupados en intervalos de clase. PMO4. Cálculo a partir de un gráfico.

c) Concepto-definición (CD). a) Media (M): CDM1 [concepto-definición asociado a la media, uno]. La definición de la media como la suma ponderada de cada uno de los valores de la variable, multiplicado por su frecuencia. CDM2. Definición de media, como promedio aritmético de un conjunto de datos. b) Mediana (ME): CDME1. Mediana como centro de la distribución. CDME2. La mediana es el valor de la variable estadística que divide en dos efectivos iguales. CDME3. El valor de la variable estadística tal que la ordenada del diagrama acumulativo de frecuencias absolutas es igual a n/2. CDME4. La mediana como el valor de la variable estadística tal que la ordenada de la curva de distribución empírica es igual a 1/2. c) Moda (MO): CDMO1. La moda es el valor más frecuente de la variable estadística. CDMO2. La moda es el valor que corresponde al máximo del diagrama de barras o histograma.

d) Propiedades (P). a) Numéricas (N): PN1 [propiedad numérica, uno]. La media, la mediana y la moda de un conjunto de datos son siempre valores pertenecientes al rango de la variable. PN2. La mediana y la media pueden no coincidir con ninguno de los valores de los datos, mientras que la moda siempre es uno de estos valores. PN3. En el cálculo de la media y la moda intervienen todos los valores de los datos, no así en el caso de la mediana. PN4. La moda y la mediana son, en ocasiones, invariantes si cambian algunos de los datos, mientras que la media sí se ve afectada por cualquier cambio en los datos. b) Algebraicas (A): PA1. Operación interna. Esta propiedad solo es válida para la moda, debido a que conserva su valor en el conjunto de datos, mientras que el cálculo de la media y la mediana no, ya que pueden tomar un valor distinto a todos sus elementos. PA2. La media, mediana y moda, consideradas como operación algebraica, no tienen elemento neutro ni simétrico. PA3. No tienen la propiedad asociativa. PA4. Son conmutativas. PA5. Las medidas de tendencia central conservan los cambios de origen y escala. PA6. La media de la suma de dos o más variables es la suma de las medias. En el caso de la mediana y la moda no se cumple. PA7. La moda puede no existir o, si existe, no ser única, mientras que la media y la mediana siempre existen. c) Estadísticas (E): PE1. La media, mediana y moda son representantes de un colectivo. PE2. La media coincide con el centro de gravedad del conjunto de datos. PE3. En distribuciones simétricas, la mediana, la media y la moda coinciden. PE4. La media es un estadístico menos resistente que la mediana y la moda. PE5. La suma de las desviaciones de un conjunto de datos con respecto a su media es cero. PE6. Es respecto a la media cuando la suma de los cuadrados de las desviaciones es mínima. PE7. Es preferible a la mediana en distribuciones con datos agrupados en los que al menos un intervalo es abierto. PE8. Existe moda/s para variables cuantitativas y cualitativas. PE9. En distribuciones con más de una moda, la mediana es el mejor representante del conjunto de datos.

e) Elementos lingüísticos. Términos, símbolos, tablas y gráficos.

f) Argumentos (ARG). ARG1 [argumento, uno]. Comprobación de casos particulares y contraejemplos. ARG2. Uso de gráficos como justificación. ARG3. Razonamientos algebraicos. ARG4. Razonamientos verbales deductivos.

5. Resultados

A continuación, se presentan los resultados del análisis de los elementos del significado de la media, mediana y moda presentes en los textos escolares chilenos de séptimo básico. Para esto, se muestran algunas imágenes para ejemplificar los elementos encontrados, se incluyen tablas que resumen los resultados, y se obtienen conclusiones sobre los hallazgos.

5.1 Situaciones-problema

En primer lugar, se analizan las situaciones-problema asociadas a cada MTC. La SPM3 se observa en la actividad de la Figura 1, donde se solicita obtener la media de las estaturas de cada conjunto de datos cuya distribución es aproximadamente simétrica.

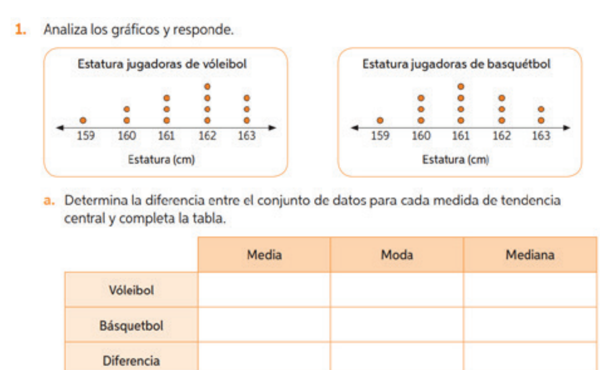


Figura 1. Situación-problema 3 asociada a la media

Nota. CA, p. 114.

En el inciso d) de la actividad de la Figura 2 se presenta la SPM5, donde la media aparece como MTC adecuada para comparar los dos conjuntos de calificaciones, en condiciones similares.

6. A continuación, se presentan las calificaciones de Alan en dos asignaturas.

Inglés: 5,6 - 5,7 - 6,0 - 6,1 - 5,8 - 6,2 - 6,2

Ciencias Naturales: 4,3 - 5,7 - 6,3 - 6,5 - 7,0 - 5,9 - 5,9

a. Calcula el rango de cada asignatura. ¿Qué puedes decir de los valores obtenidos?
b. Calcula el promedio de Alan en ambas asignaturas. ¿Qué significan esos valores?
c. Construye un gráfico de líneas en el que muestres el rendimiento de Alan en ambas asignaturas.
d. Dibuja la media en el gráfico. ¿En qué asignatura crees que la media es más representativa? Explica.

Figura 2. Situación-problema 5 asociada a la media

Nota. TE, p. 201.

En este estudio hemos considerado una nueva situación-problema asociada a la media, SPM6, donde el cálculo de la media aparece como tarea de tipo procedimental; esta se observa en la actividad de la Figura 3.

3. Calcula la media aritmética y el rango de los siguientes datos:

a. 158, 160, 168, 156, 166, 158, 160, 168, 160, 168, 158, 156, 164, 162, 166, 164, 168, 160, 162, 162, 158, 156, 166, 160, 168.

Figura 3. Situación-problema 6 asociada a la media

Nota. CA, p. 111.

En el inciso b) de la actividad de la Figura 4 se observa la SPME1, donde la mediana se presenta como la MTC más apropiada, ya que no se vería afectada por agregar un dato con valor extremo.

1. Para unas olimpiadas de Matemática los profesores han escogido 15 jóvenes que representarán al colegio. Las edades de los competidores son:

14, 11, 10, 15, 12, 15, 10, 16, 10, 10, 11, 14, 15, 16, 10.

¿Qué edad tiene como máximo la mitad más joven de los competidores?

a. Ordena los datos en forma creciente y determina el término central.
b. La respuesta anterior corresponde a la mediana. ¿Qué sucedería con la mediana si al grupo se agrega un joven de 15 años?

Figura 4. Situación-problema 1 asociada a la mediana

Nota. TE, p. 204.

La SPME3 se observa en la Figura 5, donde la mediana es la MTC más significativa para comparar los datos de cada grupo de edades.

1. Analiza el paso a paso y observa el ejercicio resuelto. Luego, realiza las actividades propuestas.

El departamento de deportes de un municipio inició las inscripciones para el taller de zumba. Cada inscrito tiene derecho a asistir a una clase semanal, las cuales se imparten los lunes (grupo 1) o los miércoles (grupo 2). Para conocer el impacto de este taller, la profesora registró las edades (años) de los asistentes durante la primera semana.

¿Cómo es la distribución de las edades en cada uno de los grupos?

Paso 1: Identifica las características de los datos de cada grupo.

Día	Asistentes	Edad promedio	Mediana	Edad mínima	Edad máxima	Rango
Lunes	11	25	16	14	45	31
Miércoles	11	25	27	21	28	7

Figura 5. Situación-problema 3 asociada a la mediana Nota. TE, p. 201.

Nota. TE, p. 206.

En este estudio hemos considerado una nueva situación-problema asociada a la mediana, SPME4, donde el cálculo de la mediana aparece como tarea de tipo procedimental; esta se observa en los tres incisos de la actividad de la Figura 6.

1. Calcula la mediana de cada conjunto de datos.

a. 7, 8, 8, 3, 2, 6, 2, 3, 6, 8, 9.
b. 2, 1, 3, 4, 3, 7, 3, 6, 5, 6, 6, 5, 4.
c. 5, 10, 15, 15, 20, 5, 5, 15, 20, 15, 25, 10.

Figura 6. Situación-problema 4 asociada a la mediana

Nota. CA, p. 113.

También, en este estudio hemos considerado una nueva situación-problema asociada a la mediana, SPME5, donde se solicita obtener la mediana para variables cuantitativas; es decir, encontrar un resumen estadístico de posición central para datos numéricos. Esta situación-problema se presenta en la actividad de la Figura 1.

En la Figura 1 se observa la SPMO1 donde se solicita la moda, es decir, el valor más frecuente de las estaturas de cada conjunto de datos. La SPMO2 se presenta en el inciso a) de la Figura 7, donde se solicita identificar la moda de un conjunto de datos cualitativos nominales.

3. Define la moda y lo que significa en cada caso.

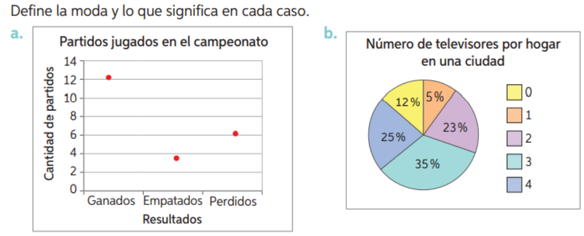


Figura 7. Situación-problema 2 asociada a la moda

Nota. TE, p. 203.

Asimismo, en este estudio hemos considerado una nueva situación-problema asociada a la moda, SPMO3, donde se solicita obtener la moda de un listado de datos como tarea de tipo procedimental (ver Figura 8).

1. Determina la moda de cada conjunto de datos.

a. 1, 2, 1, 1, 1, 5, 3, 4, 5, 4, 5, 4, 5, 6, 5, 4, 1, 2, 1, 5, 5, 6, 8.

Figura 8. Situación-problema 3 asociada a la moda

Nota. CA, p. 112.

En la Tabla 3 se presenta un resumen de las situaciones-problema que se presentan en los textos escolares chilenos de séptimo básico, considerando como referencia aquellas que se citan en Cobo (2003) y Mayén (2009). Además, añadimos aquellas (letras cursivas) que surgen del análisis propio de este artículo. Como se puede observar, no se encontraron tres situaciones-problema asociadas a la media: SPM1, SPM2 y SPM4.

Tabla 3. Situaciones-problema que presentan los textos escolares de séptimo básico. Nota. Elaboración propia.

Situación-problema	**TE**	**CA**
SPM1. Media, estimar una medida a partir de diversas mediciones realizadas, en presencia de errores		
SPM2. Media, obtener una cantidad equitativa al hacer un reparto		
SPM3. Media, obtener un elemento representativo en distribuciones aproximadamente simétricas	X	X
SPM4. Media, como el valor más probable al tomar un elemento al azar de un conjunto de valores		
SPM5. Media, para comparar dos distribuciones de datos con variables numéricas	X	X
SPM6. Media, obtener la media de un listado de datos como ejercitación		X
SPME1. Mediana, obtener la mediana cuando la media no es representativa	X	
SPME2. Mediana, obtener la posición central para variables ordinales		
SPME3. Mediana, para comparar conjuntos de datos usando, o no, gráficos	X	
SPME4. Mediana, obtener la mediana de un listado de datos como ejercitación		X
SPME5. Mediana, obtener la posición central para variables cuantitativas		X
SPMO1. Moda, obtener el valor más frecuente o dominante del conjunto de datos	X	X
SPMO2. Moda, obtener el valor representativo en datos cualitativos	X	
SPMO3. Moda, obtener la moda de un listado de datos como ejercitación		X

5.2 Procedimientos

En segundo lugar, se analizaron los procedimientos, es decir, los algoritmos de cálculo que se requieren para realizar las diversas situaciones-problema.

Con relación a los procedimientos de cálculo para la media, encontramos cuatro de los siete considerados como referencia. Estos dependen de la forma de presentación de los datos: en un listado de datos (PM1, ver Figura 2) o en un gráfico (PM4, ver Figura 1); o bien, de lo que se solicita al estudiante: obtener un dato faltante a partir del valor de la media (PM6, ver Figura 9) o determinar una distribución de datos a partir del valor de la media (PM7, ver Figura 10).

2. Si la media aritmética de 10 – 15 – 12 – X es 13, ¿Cuál es el valor de X?

A. 10
B. 12
C. 13
D. 15

Figura 9. Procedimiento 6 para la media

Nota. CA, p. 115.

4. **Desafío** Crea una tabla de frecuencia con un conjunto de 10 números que cumplan las siguientes condiciones:

a. $\overline{X} = 45$, el valor mínimo es 10, el rango es 90 y $\overline{X} >$ Me.
b. $\overline{X} = 116$, el valor máximo es 160, el rango es 70 y $\overline{X} >$ Me.

Figura 10. Procedimiento 7 para la media

Nota. TE, p. 205.

Con respecto a los procedimientos de cálculo para la mediana, encontramos cuatro de los seis considerados como referencia; estos dependen de la presentación y número de datos. El PME1 y el PME2 se presentan en los incisos a) y c) de la actividad de la Figura 6, cálculo de la mediana para datos aislados con número de datos impar y par, respectivamente. El PM3 se muestra en la actividad de la Figura 11, cálculo de la mediana a partir de los datos presentados en una tabla de frecuencias, y el PME6 se muestra en la Figura 1, cálculo de la mediana a partir de los datos presentados en un gráfico.

2. Un curso realizó una campaña solidaria para reunir azúcar: 4 estudiantes llevaron 1 kg, 11 llevaron 2 kg, 13 llevaron 3 kg y 4 aportaron con 4 kg. Identifica e interpreta la mediana.

Paso 1: Ordena los datos en una tabla de frecuencia y calcula la posición de la mediana. Como la cantidad de datos es par, entonces hay dos datos centrales, cuya posición está dada por:

$\frac{n}{2} = \frac{32}{2} = 16$ Dato central 1

$\frac{n}{2} + 1 = \frac{32}{2} + 1 = 17$ Dato central 2

Cantidad de azúcar (kg)	f	F
1	4	4
2	11	15
3	13	28
4	4	32
Total	32	

Paso 2: En F encuentra la posición 16 y 17 o que contiene a ambas. La frecuencia que las contiene es 28, por lo que la mediana es 3.

Paso 3: Interpreta la mediana. Que la mediana sea 3 kg significa que la mitad de los estudiantes que donaron la mayor cantidad de azúcar aportó como mínimo 3 kg.

Figura 11. Procedimiento 3 para la mediana

Nota. TE, p. 204.

En relación con los procedimientos de cálculo para la moda, se identificaron tres de los cuatro considerados como referencia. Estos dependen de la presentación de los datos: a) listado de datos (PMO1, ver Figura 8), tabla de frecuencias (PMO2, ver el inciso c) de la actividad de la Figura 12, donde solicita describir el cálculo de la moda a partir de la observación de los datos de una tabla) y gráfico (PMO4, ver Figura 1).

2. En un colegio ha habido distintas denuncias por *bullying cibernético*, por lo que se ha decidido investigar cuántos de los estudiantes han sido víctimas de esta práctica a través de la siguiente encuesta anónima.

Has sido víctima de *ciberbullying*?
SÍ ___ NO ___
Curso: ______

A continuación, se muestra el curso de quienes contestaron SÍ.

7° - 8° - 8° - 8° - 1° - 8° - 3° - 8° - 1° - 1° - 2° - 3° - 3°
8° - 1° - 8° - 7° - 8° - 1° - 3° - 1° - 8° - 1° - 3° - 8° - 7°

a. Ordena los datos en una tabla de frecuencia.
b. Construye un gráfico de barras con los datos.
c. ¿Cuál es la moda de los datos? Describe cómo obtenerla observando la tabla. ¿Cómo hacerlo teniendo solo el gráfico?

Figura 12. Procedimiento 2 para la moda

Nota. TE, p. 203.

En la Tabla 4 se presenta un resumen de los procedimientos de cálculo de la media, mediana y moda que se incluyen, explícita o implícitamente, en los textos escolares de nuestra muestra.

Tabla 4. Procedimientos que presentan los textos escolares de séptimo básico. Nota. Elaboración propia.

Procedimiento	**TE**	**CA**
PM1. Media, variable discreta o continua con datos aislados	X	X
PM2. Media, una variable discreta o continua con datos presentados en tablas de frecuencias		
PM3. Media, variable continua o discreta con datos agrupados en intervalos de clases		
PM4. Media, cálculo gráfico	X	X
PM5. Media, cálculo con calculadora u ordenador		
PM6. Media, inversión del algoritmo de cálculo de la media	X	X
PM7. Media, construir una distribución para una media dada	X	
PME1. Mediana, datos aislados (número de datos impar)	X	X
PME2. Mediana, datos aislados (número de datos par)	X	X
PME3. Mediana, datos en tabla de frecuencias (número de datos par)	X	
PME4. Mediana, datos en tabla de frecuencias (número de datos impar)		
PME5. Mediana, datos agrupados en clases		
PME6. Mediana, cálculo gráfico		X
PMO1. Moda, variable discreta o cualitativa con datos aislados	X	X
PMO2. Moda, variable discreta o cualitativa con datos en una tabla de frecuencias	X	
PMO3. Moda, variable discreta con datos agrupados en clases o variable continua		
PMO4. Moda, cálculo gráfico	X	X

5.3 Conceptos-definición

En tercer lugar, se analizó el concepto-definición de las MTC en la muestra de textos escolares. En el libro del estudiante encontramos expresado explícitamente un concepto-definición para cada MTC: a) CDM2, la media como promedio aritmético de un conjunto de datos (ver Figura 13); b) CDME1, la mediana como centro o valor central de una distribución de datos (ver Figura 14), y c) CDMO1, la moda como el valor más frecuente de la variable (ver Figura 15).

Se llama **media aritmética o promedio** a la cantidad total de la variable distribuida en partes iguales. La fórmula para el cálculo de esta medida de tendencia central es:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 \ldots + x_n}{n}$$

El promedio nos permite calcular un valor medio representativo de un grupo de datos, siempre y cuando el grupo no sea muy disperso, sino más bien homogéneo. La media aritmética es muy sensible a los valores que se desvían mucho del promedio.

Figura 13. Concepto-definición 2 de la media

Nota. TE, p. 199.

La **mediana** corresponde al valor que ocupa el término central de un conjunto de datos una vez ordenados de menor a mayor o viceversa. Cuando la cantidad de datos (n) de un conjunto es par, la mediana corresponde a la media aritmética de los dos términos centrales una vez que estos se ordenan.

Figura 14. Concepto-definición 1 de la mediana

Nota. TE, p. 203.

Se llama **moda (Mo)** de un conjunto de datos a la variable que presenta mayor tendencia de ocurrencia. Para calcular esta medida de tendencia central, identificamos la variable cuya frecuencia absoluta es mayor que el resto de los datos.

Un conjunto de datos puede tener más de una moda, o bien puede que no exista moda (amodal) si todos los datos se distribuyen con la misma frecuencia.

Figura 15. Concepto-definición 1 de la moda

Nota. TE, p. 202.

Además, se analizó el uso implícito que se hace de cada concepto-definición de las MTC para la solución de las situaciones-problema. A partir de esto, encontramos aquellos que no aparecen de forma explícita, sino como situaciones-problema resueltas o propuestas a los estudiantes: a) CDM1, la media como la suma ponderada de cada uno de los valores de la variable (o bien, cada uno de los valores "medios" de los intervalos de clase) multiplicado por su frecuencia, es decir, como algoritmo de cálculo (ver Figura 3); CDME2, la mediana como valor de la variable que divide en dos partes iguales a un conjunto de datos ordenados (ver paso 3 de la actividad de la Figura 11), y CDMO2, la moda como el valor que corresponde al máximo del gráfico de barras, histograma o diagrama de puntos (ver Figura 1).

En la Tabla 5 presentamos los conceptos-definición de las MTC que se han mencionado, así como la presencia de estas en cada uno de los textos escolares de séptimo básico analizados.

Tabla 5. Conceptos-definición que presentan los textos escolares de séptimo básico. Nota. Elaboración propia.

Concepto-definición	**TE**	**CA**
CDM1. Media, como algoritmo		X
CDM2. Media, como promedio	X	
CDME1. Mediana, centro de la distribución (valor central)	X	
CDME2. Mediana, divide en dos partes iguales	X	X
CDME3. Mediana en un diagrama acumulativo es n/2		
CDME4. Mediana en una distribución empírica es 1/2		
CDMO1. Moda, valor más frecuente	X	
CDMO2. Moda, valor máximo en un gráfico		X

5.4 Propiedades

Posteriormente, se analizaron las propiedades introducidas de forma explícita o implícita, clasificándolas en numéricas, algebraicas y estadísticas. Con respecto a las propiedades, reconocemos que son igual de relevantes que los conceptos-definición para el estudio y comprensión de la media, mediana y moda. No obstante, su tratamiento en los textos escolares de séptimo básico es muy inferior, ya que hemos identificado pocas: tres numéricas, una algebraica y cinco estadísticas. Por ejemplo, la PN2 (la mediana y la media pueden no coincidir con ninguno de los valores de los datos, mientras que la moda siempre es uno de estos valores) y la PN3 (para el cálculo de la media se tienen en cuenta todos valores de los datos, pero no en la mediana y la moda) se presentan, de forma implícita, en el enunciado de la Figura 16.

En este caso, la cantidad de datos sería par (16 jóvenes), por lo tanto, al ordenar los datos de menor a mayor existen dos términos centrales. En estas situaciones, cuando n es par, la mediana es la media aritmética de los dos términos centrales y no necesariamente este valor pertenece al conjunto de datos.

Figura 16. Propiedades numéricas 2 y 3

Nota. Elaboración propia.

En el libro del estudiante encontramos, en el segundo párrafo de la Figura 15, la propiedad algebraica PA7: la moda puede o no existir, mientras que la media y la mediana siempre existen. Con respecto a las propiedades estadísticas presentes en los textos escolares analizados, a continuación, ejemplificaremos tres de las cinco identificadas: PE2, la media coincide con el centro del conjunto de datos; PE3, la media, mediana y moda coinciden en distribuciones simétricas, y PE4, la media es sensible a la variación de los datos del conjunto, la mediana y la moda no, se presentan implícitamente en las Figuras 17, 18 y 13, respectivamente.

Las **medidas de dispersión** son valores numéricos que permiten medir qué tan dispersos están los datos alrededor de un valor central, por lo general la media.

Figura 17. Propiedad estadística 2

Nota. TE, p. 200.

➔ Cuando los datos sean simétricos, la media y la mediana serán esencialmente el mismo número.

Figura 18. Propiedad estadística 3

Nota. TE, p. 207.

En la Tabla 6 se presentan las propiedades numéricas, algebraicas y estadísticas de cada una de las MTC, y se señalan aquellas que se identificaron en los libros analizados.

Tabla 6. Propiedades que presentan los textos escolares de séptimo básico. Nota. Elaboración propia.

Propiedades numéricas (N), algebraicas (A) y estadísticas (E)	**TE**	**CA**
PN1. Media, mediana y moda pertenecen al rango de la variable		
PN2. La moda coincide con algún dato de la distribución, la mediana en algunos casos, pero la media no	X	X
PN3. Para la media y la moda se tienen en cuenta todos los valores del conjunto, pero no en la mediana	X	
PN4. El valor de la media cambia cuando se modifica cualquier dato del conjunto; en cambio la moda y la mediana son, en ocasiones, invariantes	X	X
PA1. La moda es una operación interna, mientras que la media y la mediana no		
PA2. La media, mediana y moda no tienen elemento neutro ni simétrico		
PA3. No tienen propiedad asociativa		
PA4. Son operaciones conmutativas		
PA5. Conservan cambios de origen y escala		
PA6. La media de la suma de dos o más variables es la suma de la media de dichas variables. En el caso de la mediana y la moda no se cumple		
PA7. La moda puede o no existir, mientras que la media y la mediana siempre existen	X	
PE1. La media, mediana y moda son representantes de un colectivo	X	
PE2. La media coincide con el centro del conjunto de datos	X	

PE3. Media, mediana y moda coinciden en distribuciones simétricas	X	
PE4. La media es sensible a la variación de los datos del conjunto, la mediana y la moda no	X	
PE5. La suma de las desviaciones del conjunto de datos con respecto a su media es cero		
PE6. Respecto a la media, la suma de cuadrados de las desviaciones es mínima		
PE7. Es preferible la mediana a la media en datos agrupados en intervalos, en los que al menos uno es abierto		
PE8. Existe moda para variables cualitativas y cuantitativas	X	
PE9. La mediana es mejor representante que la media en distribuciones no unimodales		

5.5 Elementos lingüísticos

En quinto lugar, analizamos los elementos lingüísticos. Se consideraron todos aquellos términos, palabras, expresiones, símbolos, tablas y gráficos ligados a las MTC, que sirven para enunciar los conceptos-definición y propiedades de dicho objeto, así como para describir las situaciones-problema y representar los datos.

En la Tabla 7 se presentan los elementos lingüísticos y se señalan aquellos que se identificaron en los libros analizados. En general, los elementos lingüísticos utilizados en nuestra muestra de textos escolares son similares a los resultados reportados en Cobo (2003) y Mayén (2009).

Tabla 7. Elementos lingüísticos que presentan los textos escolares de séptimo básico. Nota. Elaboración propia.

Elemento lingüístico	**TE**	**CA**
Términos	X	X
Símbolos	X	X
Tablas	X	X
Gráficos	X	X

5.6 Argumentos

Con respecto a los argumentos que se usan para justificar los procedimientos, propiedades y conceptos-definición, se encontraron ARG1 y ARG4 (ver Tabla 8), referentes a la justificación de ejemplos o contraejemplos, y razonamientos verbales deductivos, respectivamente. Estos se muestran en la Figura 11 que corresponde a un ejemplo de la mediana, en donde se explica su resolución y se presenta un argumento que justifica su resultado y su interpretación.

Tabla 8. Argumentos que presentan los textos escolares de séptimo básico. Nota. Elaboración propia.

Argumento	**TE**	**CA**
ARG1. Comprobación de casos particulares y contraejemplos	X	
ARG2. Uso de gráficos como justificación		
ARG3. Razonamiento algebraico		
ARG4. Razonamiento verbal deductivo	X	

6. Conclusiones

Las medidas de tendencia central (media, mediana y moda) son generalmente empleadas para resumir un conjunto de datos, he ahí su importancia como parte de la alfabetización estadística de cualquier ciudadano del mundo, pues estas permiten generar conclusiones, inferencias, entre otros. Así que, en este estudio, al mirar las actividades presentes en los textos escolares chilenos de séptimo básico, asumíamos que este documento representa la visión que se desarrolla alrededor de estas nociones en dicho grado escolar, la cual describiremos enseguida. Para ello, se miraron los seis elementos que describen el significado de un objeto matemático, a saber, situaciones-problema, procedimientos, conceptos-definición, propiedades, elementos lingüísticos y argumentos.

Uno de los elementos que más nos llamó la atención es que únicamente tres situaciones-problema asociadas a las MTC se comparten en ambos textos escolares: a) SPM3, la media para obtener un elemento representativo en distribuciones aproximadamente simétricas; b) SPM5, la media para comparar dos distribuciones de datos con variables numérica, y c) SPMO, la moda para obtener el valor más frecuente o dominante del conjunto de datos. Mientras, las otras situaciones-problema no se presentan (por ejemplo: SPM1, la media para estimar una medida a partir de diversas mediciones realizadas, en presencia de errores; SPM2, la media para obtener una cantidad equitativa al hacer un reparto, y SPM4, la media como el valor más probable al tomar un elemento al azar de un conjunto de valores), o bien, solo se encuentran presentes en uno de los textos escolares (por ejemplo: SPM6, obtener la media de un listado de datos como ejercitación, y SPME1, obtener la mediana cuando la media no es representativa). Esto lo percibimos como una dificultad u obstáculo para la comprensión significativa de las MTC. En comparación con estudios previos, SPM1 y SMP2 son poco usuales en los libros de textos españoles (Cobo, 2003) y mexicanos (Mayén, 2009); mientras que SPM4 tampoco se presenta en dichos estudios.

En función a los procedimientos de cálculo de las MTC que se presentan en los textos escolares analizados, identificamos que se promueven la mayoría de estos, a excepción de PM2, asociado al cálculo de la media de una variable discreta o continua con datos presentados en tablas de frecuencias; PM3, asociado al cálculo de la media de una variable continua o discreta con datos agrupados en intervalos de clases; PM5, asociado al cálculo de la media mediante el uso de una calculadora u ordenador; PME4, asociado al cálculo de la mediana de datos en tabla de frecuencias (número de datos impar); PME5, asociado al cálculo de la mediana de datos agrupados en clases; y PMO3, asociado al cálculo de la moda de una variable discreta con datos agrupados en clases o variable continua. En relación con los estudios previos, nuestros resultados coinciden en la carencia de PME4 con los textos escolares españoles (Cobo, 2003) y de PM5 y PME5 con los mexicanos (Mayén, 2009).

Con respecto a los conceptos-definición identificados en este estudio, tres se presentan de forma explícita en el texto del estudiante (CDM2, la media como promedio; CDME1, la mediana como centro de la distribución o valor central, y CDMO1, la moda como valor más frecuente) y tres de manera implícita en el cuaderno de actividades (CDM1, la media como algoritmo; CDME2, la mediana divide en dos partes iguales al conjunto de datos, y CDMO2, la moda como valor máximo en un gráfico). Estos conceptos-definición son semejantes a los encontrados en el estudio de Cobo (2003) y Mayén (2009), la única diferencia con los libros de texto mexicanos es que encontramos aquel que indica obtener la moda a partir de la lectura de un gráfico estadístico.

Dentro de las propiedades se identificó que en los textos escolares solo se presentan tres numéricas (PN2, la moda coincide con algún dato de la distribución, la mediana en algunos casos, pero media no; PN3, para la media y moda se tienen en cuenta todos los valores del conjunto, pero no en mediana; y PN4, el valor de la media cambia cuando se modifica cualquier dato del conjunto, en cambio la moda y la mediana son, en ocasiones, invariantes), una algebraica (PA7, la moda puede o no existir, mientras que la media y la mediana siempre existen) y cinco estadísticas (PE1, la media, mediana y moda son representantes de un colectivo; PE2, la media coincide con el centro del conjunto de datos; PE3, la media, mediana y moda coinciden en distribuciones simétricas; PE4, la media es sensible a la variación de los datos del conjunto, la mediana y la moda no, y PE8, existe moda para variables cualitativas y cuantitativas). Estas propiedades se presentan en mayor medida en los textos españoles (Cobo, 2003) que en los mexicanos (Mayén, 2009) analizados en estudios previos.

En relación con los elementos lingüísticos, estos muestran los cuatro tipos: términos, símbolos, tablas y gráficos. En general, estos elementos también se presentan en los resultados de los estudios de Cobo (2003) y Mayén (2009). En el caso de los argumentos asociados a las MTC, se presentan dos: comprobación de casos particulares y contraejemplos, y razonamientos verbales deductivos; mismos que se presentan en la mayoría de los textos españoles y mexicanos analizados por Cobo (2003) y Mayén (2009), respectivamente.

En la Figura 19 se presenta un hexágono que representa el significado ideal (puntos amarillos) apoyado en las seis categorías propuestas para su análisis. Los hexágonos de líneas continuas (azul y anaranjado) corresponden a la proporción del total de categorías para cada elemento de significado encontrada en el texto del estudiante (TE) y el cuaderno de actividades (CA), respectivamente. Como se puede evidenciar en esta representación, los textos escolares de séptimo básico se encuentran carentes de algunos elementos; esto implica que la comprensión que el estudiante obtenga de las MTC se encuentre posiblemente limitada.

En concreto, principalmente hay una pobreza con respecto al tipo de situaciones-problema y conceptos-definición, además de que ambos textos no comparten o promueven los mismos elementos, ni con la misma proporción; esto por las características propias de cada texto escolar. En general, la forma en la que se presentan las actividades en los libros no propicia una comprensión apropiada de las MTC, siendo solo el uso de elementos lingüísticos que se presenta desarrollado en su totalidad.

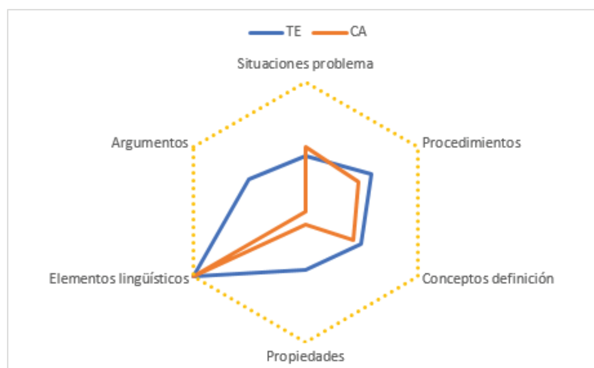


Figura 19. Distribución general de los elementos del significado de las MTC . Nota. Elaboración propia.

Con este estudio, a partir del análisis de la presencia de cada una de las categorías asociadas a cada uno de los elementos del significado de la media, mediana y moda en los textos escolares de séptimo básico, se muestra una aproximación a la forma en que estas se trabajan en las aulas chilenas. Con esta información, se recomienda diseñar una propuesta didáctica donde se aborden las diversas situaciones-problema, procedimientos, conceptos-definición, propiedades, elementos lingüísticos y argumentos de las MTC, que consideramos necesarios para su incorporación en las aulas para lograr su enseñanza más significativa. En concreto, se sugiere contextualizar los diferentes elementos de significado en actividades relacionadas con estudiantes de séptimo básico en orden de dificultad progresiva con base en el tiempo disponible.

Agradecimientos

Este trabajo ha sido desarrollado en el marco del Proyecto Fondecyt 1200005 financiado por la Agencia Nacional de Investigación y Desarrollo (ANID) de Chile.

Referencias

Batanero, C. (2000). Significado y comprensión de las medidas de posición central. *Uno. Revista de Didáctica de las Matemáticas, 25*, 41-58. https://www.ugr.es/~batanero/pages/ARTICULOS/isboa.pdf

Cabrera, B. (2014). Contenido de la media aritmética en los libros de texto y su influencia en la comprensión por estudiantes del primer ciclo de la Universidad Nacional Micaela Bastidas de Apurímac. En N. Rubio (Ed.), VII *Coloquio Internacional Enseñanza de las Matemáticas. Educación Matemática en contexto* (1173-1175). Pontificia Universidad Católica del Perú.

Cantoral, R., Montiel, G., y Reyes-Gasperini, D. (2015). Análisis del discurso matemático escolar en los libros de texto, una mirada desde la teoría socioepistemológica. *Avances de Investigación en Educación Matemática, 8*, 9-28. https://doi.org/10.35763/aiem.v1i8.123

Carvalho, J., y Gitirana, V. (2014). Média aritmética - uma análise das atividades do livro didáctico de matemática adotados no brasil. En P. Lestón (Ed.), *Acta Latinoamericana de Matemática Educativa 27* (pp. 681-688). Comité Latinoamericano de Matemática Educativa.

Chevallard, Y. (1991). *La transposición didáctica. Del saber sabio al saber enseñado*. Aique.

Cobo, B. (2003). *Significados de las medidas de posición central para los estudiantes de secundaria* [Tesis doctoral, Universidad de Granada]. https://www.ugr.es/~batanero/pages/ARTICULOS/tesiscobo.pdf

Cobo, B., y Batanero, C. (2004). Significado de la media en los libros de texto de secundaria. *Enseñanza de las Ciencias, 22*(1), 5-18. https://doi.org/10.5565/rev/ensciencias.3899

Del Pino, G., y Estrella, S. (2012). Educación estadística: relaciones con la matemática. *Pensamiento Educativo, 49*(1), 53-64. http://ojs.uc.cl/index.php/pel/article/view/25747/20671

Del Puerto, S., Seminara, S., y Minnaard, C. (2007). Identificación y análisis de los errores cometidos por los alumnos en Estadística Descriptiva. *Revista iberoamericana de educación, 43*(3), 1-8. https://rieoei.org/RIE/article/view/2331

Díaz-Levicoy, D., Morales-García, L., y Rodríguez-Alveal, F. (2020). Las medidas de tendencia central en libros de texto de Educación Primaria en México. *Revista Paradigma, 41*, 706-729. https://doi.org/10.37618/PARADIGMA.1011-2251.2020.p706-729.id819

Escobedo, J. M., y Mayén, S. (2018). Evolución en la comprensión de estudiantes de telebachillerato de un problema de estimación de media y mediana a partir de un gráfico. En J. D. Zacarías, H. Cruz, F. Velasco, B. Juárez, V. H. Vázquez, H. Reyes y F. Tajonar (Eds.), *Actualidad en la Educación Estadística y Probabilística* (pp. 111-137). Benemérita Universidad Autónoma de Puebla.

Estrella, S. (2008). Medidas de tendencia central en la enseñanza básica en Chile. *Revista Chilena de Educación Matemática, 4*(1), 20-32. http://static.ima.ucv.cl.s3.amazonaws.com/wp-content/uploads/2015/05/RECHIEM_TD_Estrella-2008-con-tapa.pdf

Estrella, S. (2016). Comprensión de la media por profesores de educación primaria en formación continua. *Revista electrónica de investigación educativa, 18*(1), 13-22. https://redie.uabc.mx/redie/article/view/635

Font, V., Godino, J. D., y D'Amore, B. (2007). An ontosemiotic approach to representations in mathematics education. *For the Learning of Mathematics, 27*(2), 2-7. http://www.dm.unibo.it/rsddm/it/articoli/damore/617%20%202007%20FLM.pdf

Font, V., Godino, J. D., y Gallardo, J. (2013). The emergence of objects from mathematical practices. *Educational Studies in Mathematics, 82*, 97-124. https://doi.org/10.1007/s10649-012-9411-0

Gal, I. (2004). Statistical Literacy, Meanings, Components, Responsibilities. En D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 47-78). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2278-6_3

Godino, J. D., y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques, 14*(3), 325-355. https://www.ugr.es/~jgodino/funciones-semioticas/03_SignificadosIP_RDM94.pdf

Godino, J. D., Batanero, C., y Font, V. (2007). The onto-semiotic approach to research in mathematics education. ZDM. *The International Journal on Mathematics Education, 39*(1-2), 127-135. https://doi.org/10.1007/s11858-006-0004-1

Godino, J. D., Batanero, C., y Font, V. (2020). El enfoque ontosemiótico: implicaciones sobre el carácter prescriptivo de la didáctica. *Revista Chilena De Educación Matemática, 12(*2), 47-59. https://doi.org/10.46219/rechiem.v12i2.25

Herbel, B. A. (2007). From intended curriculum to written curriculum: Examining the "voice" of a mathematics textbook. *Journal for Research in Mathematics Education, 38*(4), 344-369. https://www.jstor.org/stable/30034878?seq=1

Hernández, R., Fernández, C., y Baptista, P. (2010). *Metodología de la investigación.* McGraw Hill.

López-Noguero, F. (2002). El análisis de contenido como método de investigación. XXI. *Revista de Educación, 4*, 167-180. http://rabida.uhu.es/dspace/handle/10272/1912

Mayén, S. (2009). *Comprensión de las medidas de tendencia central por estudiantes mexicanos de Educación Secundaria y Bachillerato* [Tesis doctoral, Universidad de Granada]. Repositorio Institucional de la Universidad de Granada. https://digibug.ugr.es/handle/10481/2418

Mayén, S., Cobo, B., Batanero, C., y Balderas, P. (2007). Comprensión de las medidas de posición central en estudiantes mexicanos de bachillerato. *Unión. Revista Iberoamericana de Educación Matemática, 9*, 187-201. http://www.fisem.org/www/union/revistas/2007/9/Union_009_016.pdf

Ministerio de Educación de Chile. (2016). *Bases Curriculares. 7° básico a 2°* medio. Autor.

Ministerio de Educación de Chile. (2018). *Bases Curriculares. Primero a Sexto Básico*. Autor.

Ocoró, L. V., y Ocoró, S. (2016). Análisis de las medidas de tendencia central en dos libros de textos escolares de grado séptimo: el caso de la media aritmética. En I. Álvarez y C. Sua (Eds.), *Memorias del II Encuentro Colombiano de Educación Estocástica* (184-190). Asociación Colombiana de Educación Estocástica.

Pérez-Serrano, G. (1994). *Investigación cualitativa. Retos e interrogantes. I. Métodos*. La Muralla.

Pino-Fan, L. (2013). *Evaluación de la faceta epistémica del conocimiento didáctico-matemático de futuros profesores de bachillerato sobre la derivada* [Tesis doctoral, Universidad de Granada]. https://www.ugr.es/~jgodino/Tesis_doctorales/Luis_Pino_tesis.pdf

Vásquez, C., y Alsina, Á. (2014). Enseñanza de la Probabilidad en educación primaria. Un desafío para la formación inicial y continua del profesorado. *Números. Revista de Didáctica de las Matemáticas, 85,* 5-23. http://www.sinewton.org/numeros/numeros/85/Articulos_01.pdf

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

ARTÍCULOS DE INVESTIGACIÓN

https://doi.org/10.46219/rechiem.v13i4.83

PREGUNTAS ELABORADAS POR PROFESORES PARA EL ESTUDIO DE GRÁFICOS DE BARRAS ESTADÍSTICOS: LOS NIVELES DE LECTURA QUE SE IDENTIFICAN EN SUS PROPUESTAS

QUESTIONS ADDRESSED BY TEACHERS FOR THE STUDY OF STATISTICAL BAR GRAPHICS: THE READING LEVELS THAT ARE IDENTIFIED IN THEIR PROPOSALS

Fabiola Arévalo-Meneses
fabiola.arevalo.m@mail.pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Julio Manzanares
julio.manzanares.c@mail.pucv.cl
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

RESUMEN

Esta investigación de enfoque cualitativo expone los resultados de un estudio exploratorio cuyo objetivo fue mostrar el nivel de lectura de gráficos estadísticos presentes en las preguntas propuestas por docentes de Educación Matemática en Chile, ante gráficos que presentan datos de contextos contingentes en un periodo de tiempo afectado por la situación sanitaria por COVID-19 y datos que carecen de contexto. Los datos obtenidos fueron categorizados de acuerdo a los niveles de lectura de gráficos estadísticos, en donde observamos una tendencia a la elaboración de preguntas que abordan niveles básicos de lectura de gráficos estadísticos, analizando si estos tienen relación con la formación inicial de los profesores, su experiencia docente y si existe diferencia ante los gráficos presentados. Se hace relevante incorporar datos de temáticas contingentes en el estudio de gráficos estadísticos, pues promueven un nivel de lectura mayor de los datos, lo que reafirma la importancia que adquiere el contexto en el estudio estadístico, así como la necesidad de fortalecer el desarrollo profesional docente en torno a la alfabetización estadística, con el fin de formar ciudadanos críticos en una sociedad en constante cambio.

PALABRAS CLAVE:
Educación estadística, Formación de profesores, Gráficos estadísticos, Educación en contextos contingentes.

ABSTRACT

This research with a qualitative approach shows the results of an exploratory study whose objective was to show the reading level of statistical graphs present in the questions addressed by mathematics teachers in Chile. The graphs considered present data from contingent contexts in a period affected by the health situation of COVID-19 and the lack of context data. The data obtained were categorized according to the reading levels of statistical graphs where we observed a trend towards the elaboration of questions that address basic reading levels of statistical graphs, analyzing whether these are related to the initial formation of the teachers, their teaching experience and if there is a difference from the graphs presented. It is relevant to incorporate data on contingent topics in the study of statistical graphics, since they promote a higher reading level of the data, which reaffirms the importance of context in statistical studies, as well as the need to strengthen teacher professional development around statistical literacy, in order to train critical citizens in an ever changing society.

KEYWORDS:
Statistical education, Teacher training, Statistical graphics, Education in contingent contexts.

Recibido: 30 de Mayo de 2021, Aceptado: 19 de Julio de 2021

1. Introducción

Actualmente nos encontramos en un escenario complejo producto de la pandemia por COVID-19, que ha afectado en gran medida la vida de las personas. En este contexto, los distintos medios de comunicación nos están entregando constantemente mucha información, ante la cual, como receptores, de manera espontánea, adquirimos una visión y una percepción específica, nos formamos ideas, opiniones, juicios de valor y una postura ante la temática en cuestión. De esta manera, cobra especial importancia la interpretación y validación de información, interpelándonos como ciudadanos.

Desde la educación escolar, la formación de futuros ciudadanos capaces de insertarse en una sociedad en constante cambio resulta indispensable en el desarrollo integral de los estudiantes. Es en este escenario que la alfabetización estadística juega un rol fundamental, idea que es considerada en los planes y programas propuestos por el Ministerio de Educación de Chile (MINEDUC), los que se orientan al desarrollo de habilidades, actitudes y conocimientos en nuestros estudiantes, considerando objetivos que se enfocan en la formación integral de estos como futuros ciudadanos críticos. En Educación Matemática estos objetivos son considerados en el eje temático de Probabilidades y Estadística.

> Este eje responde a la necesidad de que todos los y todas las estudiantes aprendan a realizar análisis, inferencias y obtengan información a partir de datos estadísticos. Se espera formar alumnos críticos y alumnas críticas que puedan utilizar la información para validar sus opiniones y decisiones; que sean capaces de determinar situaciones conflictivas a raíz de interpretaciones erróneas de un gráfico y de las posibles manipulaciones intencionadas que se pueden hacer con los datos. (MINEDUC, 2016, p. 102)

El gráfico estadístico es un instrumento fundamental para la organización y presentación de datos, los que se pueden describir y analizar, permitiéndonos obtener y comunicar información (Arteaga et al., 2011). Esto revela la importancia que adquiere el desarrollo de habilidades y competencias en torno al estudio de gráficos estadísticos en la educación escolar. En consecuencia, nos podemos preguntar ¿cómo se abordan los gráficos estadísticos en el aula? los profesores, ¿poseen la formación estadística que les permita abordar de manera óptima esta tarea?

Existen múltiples investigaciones en el ámbito de la educación estadística enfocadas en la lectura e interpretación de gráficos estadísticos (Aoyama y Stephens, 2003; Bertin, 1967; Curcio, 1987; Friel et al., 2001; Mautone y Mayer, 2007; Shaughnessy et al., 1996). En Contreras et al. (2017), mediante un cuestionario aplicado a futuros profesores de primaria orientado a evaluar aspectos relevantes de la cultura estadística, se observó que existe una escasa comprensión gráfica, concluyendo acerca de la importancia que adquiere la revisión de los programas de formación de profesores. En el mismo sentido, Espinel (2007) muestra resultados de dos investigaciones anteriores, también en futuros profesores de primaria, una de ellas referente a construcción y evaluación de histogramas y otra sobre representaciones de distribución de datos, donde se observan los errores frecuentes que cometen en la construcción de gráficas y dificultades que presentan en el razonamiento frente a distintas representaciones de distribuciones de datos. Por otra parte, en Arteaga et al. (2011) se reúnen los aspectos más importantes de investigaciones referentes a las competencias que se requieren para la construcción y lectura de tablas y gráficos estadísticos, destacando la necesidad de que los profesores estén bien preparados para así formar ciudadanos estadísticamente cultos, lo que se puede lograr mejorando la cultura estadística en futuros profesores a través de su formación. Otra investigación realizada por Estrella et al. (2015) considera tres elementos fundamentales dentro de la educación estadística, entre los cuales se encuentra la comprensión gráfica. En este estudio se aplicó un cuestionario a profesores de educación primaria en Chile y a sus estudiantes, cuyos resultados mostraron un muy bajo desempeño en la lectura interpretativa de gráficos estadísticos tanto en profesores como en estudiantes, haciendo énfasis en la necesidad de que los profesores se enfrenten a situaciones de incerteza en el estudio estadístico.

Rodríguez y Sandoval (2012) indagaron en el nivel de lectura de gráficos estadísticos en 47 profesores de primaria y 44 futuros profesores de primaria en Chile, comparando ambos grupos en relación a la habilidad de codificación y decodificación de información gráfica, evidenciando un bajo nivel tanto en profesores como en futuros profesores de primaria.

Por su parte, Díaz-Levicoy et al. (2015) analizaron los tipos de gráficos y sus niveles de lectura presentes en 12 textos de educación primaria en Chile, en donde la mayoría eran gráficos de barras y pictogramas, siendo "leer dentro de los datos" el nivel de lectura más frecuente. Es así que se observa la necesidad de incorporar actividades que contemplen niveles de lectura de datos más profundos en cursos superiores. De esta forma, surge de manera natural la preocupación de indagar en relación a cómo se enseña estadística en la educación escolar, cuáles son las limitaciones que presenta el docente al momento de enseñar estadística, y puntualmente cómo aborda el estudio de gráficos estadístico con sus estudiantes, qué habilidades y competencias espera desarrollar en ellos. Sin dudas es una situación interesante de ser investigada, pues es a través de la alfabetización estadística que los estudiantes podrán ser ciudadanos efectivos en la sociedad de la información (Pino y Estrella, 2012).

Considerando el contexto de pandemia producto de COVID-19, se han llevado a cabo varias investigaciones desde esta temática. Alsina et al. (2020) describen experiencias de aprendizaje relacionadas con la pandemia, tomando como referencia investigaciones anteriores que declaran la importancia de considerar temáticas contextuales que aporten significado al estudio estadístico. Del mismo modo, Coob y Moore (1997) señalan que los datos son números que se encuentran en un contexto, lo que les da significado. Por otra parte, un estudio reciente (Vásquez, 2021), en el mismo marco de la emergencia sanitaria por COVID-19, analizó las preguntas que plantearon 48 futuros profesores de educación primaria en Chile para la comprensión de gráficos y el desarrollo de competencias de sostenibilidad en sus estudiantes. Los resultados mostraron que la mayoría de las preguntas planteadas por los futuros profesores se encuentran en los niveles 1 y 2 ("leer los datos" y "leer dentro de los datos"), que se refieren a preguntas que no requieren de un análisis profundo de los datos, indicando que puede deberse a que las preguntas elaboradas por futuros profesores van dirigidas a estudiantes de primaria, motivo por el cual el nivel de dificultad no puede ser mayor, o bien que los futuros profesores participantes no son capaces de plantear preguntas que consideren niveles de lectura más profundos. Por otra parte, señalan que solo dos de los profesores plantearon preguntas características del nivel 4 "leer detrás de los datos", que son aquellas que contribuyen al desarrollo del pensamiento crítico y comprensión del contexto en que se encuentra la información.

Si bien el presente estudio se desarrolló sin conocer la investigación mencionada anteriormente, dada su reciente publicación, podemos observar que considera elementos comunes a los que se abordan en esta investigación, lo que nos brinda la posibilidad de poder contrastar y comparar en el análisis de los datos. Ahora bien, el presente estudio considera a profesores de educación primaria y profesores de Matemáticas, los cuales ya terminaron su formación inicial docente. Por otra parte, se consideran gráficos de barras estadísticos que, de acuerdo a los datos que presentan, no necesariamente tienen relación con la crisis sanitaria vivida actualmente, sin embargo, no se hace referencia al respecto con el fin de no interferir en la interpretación de la información por parte de los profesores.

Considerando los antecedentes expuestos, el objetivo de este estudio es indagar en los niveles de lectura de gráficos estadísticos presentes en preguntas elaboradas por profesores que actualmente realizan docencia, dirigidas a sus estudiantes para el estudio de gráficos de barras estadísticos. Y a la vez, identificar si la formación inicial y experiencia de los profesores, así como la temática relacionada con los datos presentados en el gráfico, inciden en las preguntas que realizan.

2. Marco de referencia

La base de esta investigación radica en el análisis de los niveles de lectura e interpretación de gráficos estadísticos, comprendiendo que la lectura comienza con una identificación de lo externo (barras, título, etc.), y luego lo interno (variables, significados de las variables, escalas, etc.), distinguiendo tres niveles de lectura de gráficos: N1: Extracción de los datos; N2: Extracción de las tendencias, y N3: Análisis de la estructura de los datos (Bertin, 1967).

Por otra parte, Curcio (1987) señala que para la lectura y construcción de gráficos se debe comprender tres aspectos que son característicos en ellos:

- Las expresiones o palabras.
- Lo matemático que se subyace.
- El convenio de construcción de gráficos.

Más tarde, por la necesidad de la comprensión de todos estos aspectos, Curcio (1989) estableció niveles de lectura, ampliando la definición de Bertin (1967). Estos son:

L1: Leer los datos: cuando se realiza una lectura literal de algún dato representado en el gráfico estadístico.

L2: Leer dentro de los datos: cuando se realiza una lectura de un dato del gráfico que no se encuentra de forma explícita en él, y que requiere de alguna operación matemática (adición, promedio, etc.) para encontrarlo.

L3: Leer más allá de los datos: cuando se obtiene una información que no se presenta directamente en el gráfico, y no es posible encontrarla por medio de cálculos u operaciones, infiriendo dicho dato.

Con el pasar del tiempo, y por la necesidad que se fue presentando en la investigación respecto del análisis y comprensión de los gráficos estadísticos, Friel et al. (2001) ampliaron nuevamente lo propuesto por Curcio, agregando una clasificación más. Esta es:

L4: Leer detrás de los datos: cuando se realiza una valoración crítica de la información presentada en el gráfico estadístico sobre la recogida de datos, su organización o las conclusiones. A ello se suma el conocimiento del contexto en el que se extraen los datos del gráfico.

En base a lo planteado por Curcio (1989) y Friel et al. (2001), esta investigación utiliza los niveles de lectura de gráficos estadísticos de los autores señalados, para analizar el enfoque de las preguntas realizadas por docentes de Matemática en base a gráficos estadísticos.

3. Metodología

La presente investigación se realiza mediante una metodología de tipo cualitativa (Hernández et al., 2015). A través del análisis de contenido (Krippendorff, 2013), se busca indagar en las preguntas que proponen profesores de Educación Matemática, dirigidas a sus estudiantes, para el estudio de dos gráficos de barras estadísticos.

3.1 Caracterización de los sujetos de estudio

Los sujetos de estudio son docentes de enseñanza de la Matemática que se encuentran realizando sus funciones pedagógicas en establecimientos escolares de la quinta región de Valparaíso, Chile. En total, 22 docentes respondieron el instrumento, entre los que se encuentran profesores cuya formación inicial es profesor de educación primaria, quienes realizan docencia en educación primaria. Y profesores de formación inicial de pedagogía en Matemáticas, quienes realizan docencia en educación secundaria.

3.2 Diseño del instrumento

El instrumento que se utilizó para recoger los datos fue un formulario, el cual por la contingencia que se suscita a nivel mundial (pandemia COVID-19), se tuvo que realizar de manera online, por medio de la plataforma Google (Google Forms), a través de la cual se buscó en primera instancia caracterizar a los sujetos de investigación, consultando por su formación inicial y los años de experiencia que poseen en el sistema escolar. Posteriormente se les presentó dos gráficos de barras con la finalidad de conocer qué tipo de preguntas dirigidas a sus estudiantes realizarían en relación a estos.

El primer gráfico (G1) presenta información sobre la cantidad de latas recolectadas por una comunidad, en donde se puede observar la cantidad recolectada por hombres y por mujeres en cada año, desde el año 2014 al año 2017 (ver Figura 1).

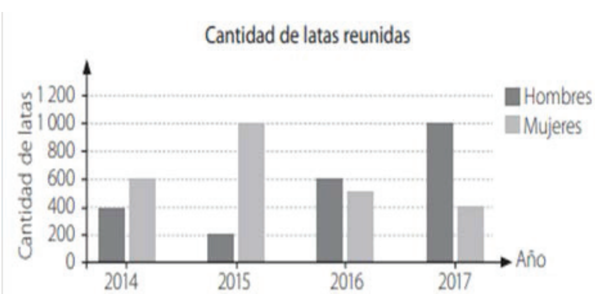


Figura 1. Gráfico (G1), cantidad de latas reunidas por una comunidad. Nota. Imagen de recursos de la unidad para 6º básico (Curriculumnacional.mineduc.cl).

El gráfico G2 corresponde a una boleta de consumo de agua de un hogar ubicado en la quinta región de Valparaíso, en que se muestra el consumo de agua desde abril de 2019 a abril de 2020 (ver Figura 2). Cabe destacar que es durante este periodo de tiempo que Chile entra en estado de emergencia sanitaria por la presencia de COVID-19, específicamente en marzo del año 2020. Ahora bien, el gráfico se muestra tal y como se puede observar en la Figura 2, es decir, no se indican las posibles causas del aumento en el consumo de los últimos meses, pudiendo deberse a diferentes factores. El propósito es no interferir en las interpretaciones y el sentido que el profesor puede dar a la información que entrega el gráfico, lo que puede verse reflejado en las preguntas que elabora para sus estudiantes.

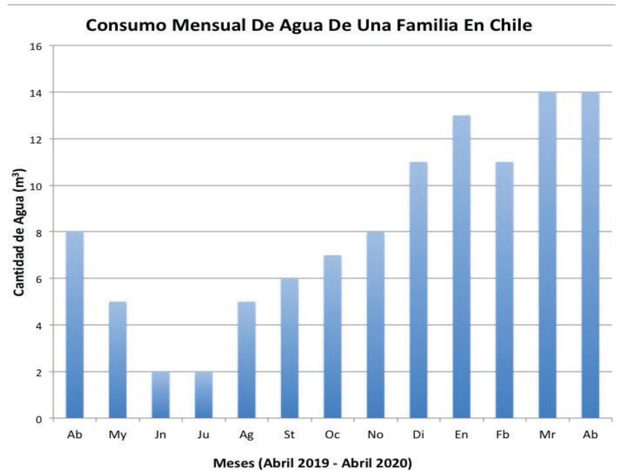


Figura 2. Gráfico 2 (G2), consumo mensual de agua de una familia en Chile. Nota. Boleta de Esval de un departamento en la V región, Chile.

Para cada uno de los gráficos (G1 y G2) se les pidió a los profesores que diseñaran preguntas que ellos plantearían a sus estudiantes para el estudio de estos.

3.3 Recogida de datos

Dadas las actuales medidas de seguridad que se han tomado en el marco de la crisis sanitaria por COVID-19, se consideró la mejor alternativa en cuanto a herramientas tecnológicas que nos permiten comunicarnos de forma efectiva y que a la vez son de fácil acceso para los profesores, de este modo, se envió un formulario vía correo electrónico. Esto permitió contar de forma inmediata con los datos una vez que los profesores respondieran el cuestionario. De esta manera se obtienen los siguientes resultados (ver Tabla 1).

Tabla 1. Caracterización de los datos obtenidos

Nota. Elaboración propia.

Cantidad de preguntas por gráficos	
G1	G2
61	55
Total de preguntas: 116	

Profesores por formación inicial	
Educación básica	Pedagogía en Matemática
6	16
Total de profesores: 22	

Profesores por experiencia laboral			
Entre 0 y 3 años	Entre 3 y 7 años	Entre 7 y 12 años	Más de 12 años
4	6	6	6
Total de profesores: 22			

3.4 Categorías de análisis

Considerando los niveles de lectura de gráficos propuestos por Friel et al. (2001), se definen en la Tabla 2 las siguientes categorías para clasificar las preguntas realizadas por los profesores en ambos gráficos propuestos, de acuerdo a la intencionalidad que se puede apreciar en las pregunta realizadas en relación a lo que motivan en el estudiante.

Tabla 2. Categoría de análisis

Nota. Elaboración propia.

Sigla	**Categoría**	**Definición**
L1	Preguntas de nivel L1 de lectura de gráficos.	La pregunta induce a la lectura literal de datos.
L2	Preguntas de nivel L2 de lectura de gráficos.	La pregunta induce a realizar una lectura de un dato del gráfico que no se encuentra de forma explícita en él, y que requiere de alguna operación matemática (adición, promedio, etc.) para encontrarlo.
L3	Preguntas de nivel L3 de lectura de gráficos.	La pregunta induce a extraer información que no se presenta directamente en el gráfico, y no es posible encontrarla por medio de cálculos u operaciones, infiriendo dicho dato.
L4	Preguntas de nivel L4 de lectura de gráficos.	La pregunta induce a realizar una valoración crítica de la información presentada en el gráfico (sobre la recogida de datos, su organización o las conclusiones). A ello se suma el conocimiento del contexto en el que se extraen los datos del gráfico.

4. Resultados

A partir de los datos obtenidos se presentan a continuación los resultados en relación a las preguntas elaboradas al comparar ambos gráficos respecto de la formación inicial y de la experiencia docente.

4.1 Resultados en relación a los gráficos presentados

La Figura 3 presenta el resultado en relación a la clasificación de las preguntas por gráficos.

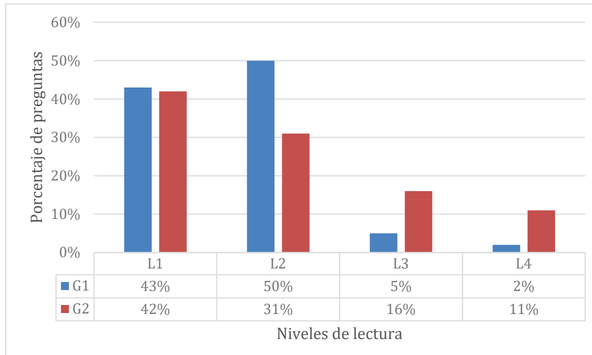


Figura 3. Clasificación de las preguntas por gráficos

Nota. Elaboración propia.

En la Figura 3 se representa la relación entre los niveles de lectura de gráfico que abordan las preguntas realizadas y el gráfico al que pertenecen. Podemos observar que la mayoría de las preguntas elaboradas por los docentes se encuentran en el nivel L1 (leer los datos) y L2 (leer dentro de los datos) de lectura de gráficos de la taxonomía de Curcio (1987), que en conjunto corresponden a un 83% de las preguntas (ver ejemplos en Tabla 3). Mientras que una cantidad significativamente menor (10%) se encuentra en el nivel L3 (leer más allá de los datos) y tan solo un 6% corresponde a preguntas del nivel L4 (leer detrás de los datos), propuesto por Friel et al. (2001).

Tabla 3. Ejemplos de preguntas de niveles L1 y L2

Nota. Preguntas extraídas de datos obtenidos.

Gráfico	**Nivel**	**Pregunta realizada por profesor (P)**
G1	L1	¿En qué año se logró la mayor recolección de latas por parte de las mujeres? (P13)
G2	L2	¿Cuál es la diferencia de consumo de agua entre los meses de abril 2019 y abril 2020? (P8)

Considerando las preguntas realizadas de acuerdo al gráfico 1 (G1) y al gráfico 2 (G2), podemos notar que en los niveles L1 y L2 se presenta una diferencia considerable, en tanto que en G1 estos niveles concentran el 93% de las preguntas, mientras que en G2, un 73%. Por otra parte, en los niveles L3 y L4 existe una gran diferencia entre los gráficos G1 (7%) y G2 (27%) (ver ejemplos en Tabla 4).

Tabla 4. Ejemplos de preguntas niveles L3 y L4

Nota. Preguntas extraídas de datos obtenidos.

Gráfico	Nivel	Pregunta realizada por profesor (P)
G1	L3	¿Qué puedes inferir respecto de la recolección de latas realizadas por las mujeres y los hombres durante los 4 años? (P21)
G2	L4	¿Qué circunstancia crees que existieron durante el periodo de mayor consumo de agua en la familia? (P15)

De esta manera, podemos ver que la mayoría de las preguntas de los niveles L1 y L2 pertenecen a G1 (recolección de latas), y la mayoría de las preguntas de los niveles L3 y L4 se encuentran en G2 (consumo de agua).

4.2 Resultados en relación a la formación inicial

La Figura 4 presenta el resultado en relación a la clasificación de las preguntas de acuerdo a la formación inicial.

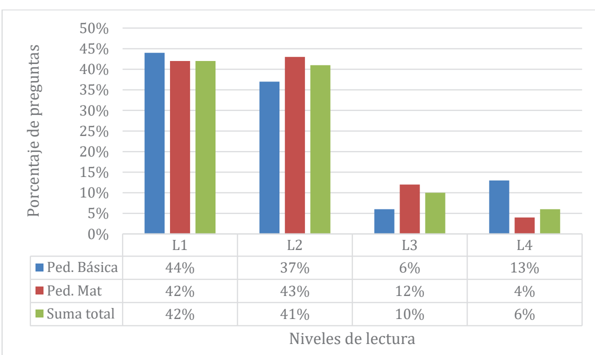


Figura 4. Clasificación de las preguntas por formación inicial. Nota. Elaboración propia.

La Figura 4 muestra la relación entre la formación inicial de los docentes y los niveles de lectura de gráficos presentes en sus preguntas, en los cuales podemos observar que no existe mayor diferencia porcentual entre la cantidad de preguntas realizadas en los niveles L1 y L2 por docentes cuya formación inicial es pedagogía en educación primaria (44% y 38%, respectivamente), y las realizadas por profesores cuya formación inicial es pedagogía en Educación Matemática (42% y 43%, respectivamente). Por otra parte, la mayor cantidad de preguntas del nivel L3 son realizadas por profesores de educación secundaria (12%), en comparación a educación primaria (6%). Ahora bien, la relación se invierte en el nivel L4, siendo mayor la cantidad de preguntas realizadas por profesores de educación primaria (13%), mientras que en educación secundaria es tan solo un 4%. Sin embargo, no podemos inferir respecto de estas diferencias en el nivel L3 y L4 en relación a la formación docente dado que la cantidad de preguntas de estos niveles fueron muy escasas.

4.3 Resultados en relación a la experiencia docente

La Figura 5 presenta el resultado en relación a la clasificación de las preguntas de acuerdo a la experiencia docente.

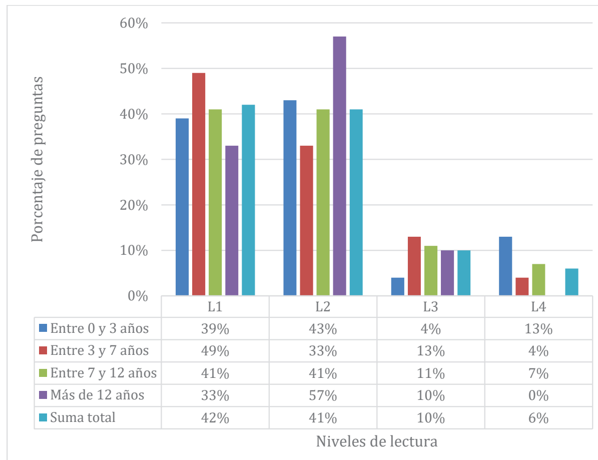


Figura 5. Clasificación de las preguntas por años de experiencia docente. Nota. Elaboración propia.

De la relación entre la experiencia laboral y el nivel de lectura de gráficos de las preguntas realizadas, se observa que en todos los rangos de años de experiencia docente se concentra al menos el 83% de las preguntas en los niveles L1 y L2, y tan solo un 16% de las preguntas se encuentran en los niveles L3 y L4, siendo en el rango de 0 a 3 años de experiencia el que presenta mayor cantidad de preguntas del nivel L4 (13%), mientras que, en este mismo nivel de lectura de gráficos, los docentes que poseen 12 o más años de experiencia (27% de los encuestados), no realizan ninguna pregunta. Así también, podemos notar en este mismo rango de experiencia, que más de la mitad de las preguntas (57%) son preguntas pertenecientes al nivel L2.

Podemos afirmar de esta manera que, en todos los rangos de experiencia docente, se mantiene la tendencia a realizar una gran cantidad de preguntas de nivel L1 y L2, que apuntan a la descripción e interpretación de los datos, mientras que, entre más años de experiencia docente tengan los profesores, la cantidad de preguntas realizadas en los niveles L3 y L4 tienden a disminuir y a desaparecer, que son aquellas que buscan realizar inferencias y una valoración crítica de los datos.

5. Análisis

De acuerdo al análisis de los datos obtenidos, constatamos que efectivamente las preguntas elaboradas por los profesores se concentran en los niveles de lectura L1 (leer los datos) y L2 (leer dentro de los datos), de acuerdo a la taxonomía de lectura de gráficos (Curcio, 1987; Friel et al., 2001). Este resultado concuerda con lo observado por Alsina et al. (2020), Contreras et al. (2017), Estrella et al. (2015) y recientemente por Vásquez (2021) en futuros profesores de educación primaria, de los cuales muy pocos alcanzan los niveles más altos en lectura de gráficos. Por otra parte, en una investigación en que se analizaron textos escolares de educación primaria en Chile y España, se observó que el nivel más frecuente es "leer dentro de los datos" (Díaz-Levicoy et al., 2015), lo que podría estar relacionado con la tendencia de los profesores a presentar preguntas enfocadas a obtener datos desde la lectura de las etiquetas y valores presentes en el gráfico así como la comparación de datos, y no a la valoración crítica de la información que se puede inferir desde el gráfico.

En relación a la experiencia y la formación inicial docente, no es posible observar que tenga relación directa con el nivel de lectura de gráficos presente en las preguntas. No obstante, un caso particular fue el de aquellos profesores con 12 o más años de experiencia, de los cuales ninguno realizó preguntas del nivel L4 "leer detrás de los datos". Es posible que dicho suceso pueda estar relacionado con las áreas de conocimiento que los profesores han abordado durante sus años de experiencia docente, pues los datos obtenidos no nos proporcionan información concreta en relación a cuánto tiempo de su vida profesional han dedicado a la enseñanza de la estadística, por lo que no podemos conjeturar al respecto. Pino y Estrella (2012) señalan que la formación académica de los docentes es débil y que sus experiencias en educación estadística son mínimas. Considerando lo anterior, resulta necesario poder indagar en la experiencia de los profesores, particularmente en enseñanza de estadística durante sus años en ejercicio, así como indagar y profundizar en la formación profesional. Sin embargo, más allá de este caso particular, de manera transversal se destaca la casi inexistencia de preguntas del nivel L4, lo que también se observa en Vásquez (2021) con futuros profesores de primaria, en donde ninguno de ellos propuso preguntas de este nivel. Sin dudas esto resulta preocupante y se transforma en un elemento que debe ser considerado al momento de diseñar programas de desarrollo profesional docente.

De acuerdo a los resultados de esta investigación, en profesores de educación básica si bien la mayoría de sus preguntas son de niveles L1 y L2, lo que está acorde a lo establecido por las bases curriculares, creemos necesario incorporar preguntas de los niveles L3 y L4 en los últimos cursos de la enseñanza básica, tal como también lo afirman Díaz-Levicoy et al. (2015). Lo que observamos en preguntas de los profesores de educación secundaria es similar a lo ocurrido en educación primaria, es decir, la mayoría de las preguntas son de niveles L1 y L2, y se alejan de lo expresado en las bases curriculares en el eje de Datos y Azar, pues en este ciclo se espera que los estudiantes realicen inferencias y valoración crítica de la información a partir de datos estadístico, niveles L3 y L4 (MINEDUC, 2016).

Por otra parte, en relación a los gráficos presentados a los docentes, notamos que existe una situación particular en el gráfico G2 (consumo de agua), pues es en este donde se concentra la mayoría de las preguntas de los niveles L3 y L4. El gráfico G2 presenta el consumo de agua de una vivienda en la ciudad de Valparaíso, en el periodo de abril de 2019 hasta abril de 2020. Desde marzo del 2020 Chile se ha visto afectado por la crisis sanitaria por COVID-19, situación que ha provocado un incremento en los consumos básicos del hogar, tal como se indica en el estudio publicado por el Centro del Agua para Zonas Áridas y Semiáridas de América Latina y el Caribe (CAZALAC) en Mansilla et al. (2020), que hace referencia a la demanda hídrica en Chile ante la pandemia. El estudio revela que el consumo de agua potable en un hogar aumenta al 69,3% en relación al consumo habitual, durante el periodo afectado por la crisis sanitaria e hídrica del país. Este aumento podría deberse a las medidas que se han tomado para prevenir el contagio de COVID-19, tales como mayor frecuencia de higiene personal, del hogar y de los alimentos. Es así, que en el gráfico G2, se presentan datos en relación a una temática de contingencia mundial, por tanto, de gran interés y preocupación, que afecta hoy en día directamente la vida de las personas.

Es importante recordar que el gráfico fue presentado a los profesores sin realizar ninguna observación en relación al contexto ni a las posibles causas del aumento en el consumo de agua, las que podrían ser múltiples y diversas. Esto nos lleva a pensar que el hecho de estar viviendo una situación tan particular como la crisis sanitaria por COVID-19, parece predisponer de manera casi innata al observador del gráfico, de tal modo que advierte el aumento en el consumo y lo relaciona directamente con lo que se vive en tiempos de pandemia. Podríamos decir que la elaboración de preguntas de los niveles L3 y L4 que promueven un análisis crítico de la información se ve motivada en parte por los conocimientos adquiridos por el profesor de acuerdo a su propia experiencia personal, desde donde da sentido a la información. De cierto modo, el profesor agrega conocimiento referente al gráfico, lo que se puede apreciar en la intencionalidad de las preguntas que elabora. Es así como el conocimiento del profesor en relación a la información y contexto de los datos que se presentan en el gráfico, resultan un factor muy importante al momento en que el profesor realiza preguntas orientadas al estudio de gráficos estadísticos. Si bien, necesitamos más datos que nos permitan conjeturar al respecto, existen estudios que destacan la relevancia que adquiere el contexto en la enseñanza de la estadística. Coob y Moore (1997), así como Pino y Estrella (2012), expresan que en la enseñanza de la estadística el contexto en que se presentan los datos es de suma importancia para la alfabetización estadística. Del mismo modo, Batanero et al. (2013) relacionan la cultura estadística y el razonamiento estadístico como las dos componentes que dan sentido al estudio estadístico, indicando que sin considerar el contexto donde se encuentran los datos, se pierde dicho sentido. Esta idea es apoyada en Alsina et al. (2020) y en Vásquez (2021), quienes en recientes investigaciones presentan una serie de experiencias contextualizadas con datos de COVID-19, en que concluyen que mediante estas actividades los estudiantes desarrollan la capacidad de reflexionar y ser críticos ante la información.

6. Conclusiones

En este estudio se observó que, independiente de la formación inicial del profesor, de la experiencia docente y de los datos que se presentan en el gráfico de estudio, la gran mayoría de las preguntas que los profesores realizan ante el estudio de gráficos estadísticos se encuentran en los nivel de lectura L1 y L2, a las que se puede dar respuesta desde la lectura literal de los datos y de interpretación de datos que se fundamentan en la comparación y cálculos a partir de los datos, siendo escasa y casi nula la propuesta de preguntas que propongan al estudiante realizar un nivel más profundo de lectura de gráficos, que apunten a realizar predicciones e inferencias, así como un análisis crítico de la información. Esto presenta un gran desafío para las instituciones de formación inicial y continua de profesores, que promueven la adquisición de conocimientos de la disciplina y su didáctica específica, con el fin de formar ciudadanos estadísticamente cultos y comprometidos socialmente (Arteaga et al., 2011).

No obstante, podemos decir que esto no es suficiente, pues hemos podido observar la importancia que adquiere el conocimiento que el profesor tenga del contexto en relación a los datos presentes en el gráfico, lo que le permite intensionar un análisis más profundo y crítico de la información. Esto presenta un desafío importante para el docente, quien debe mantenerse constantemente informado en relación a temáticas de interés colectivo y que sean significativas para sus estudiantes. En consecuencia, no se puede perder de vista lo importante que es incorporar situaciones contextualizadas, que sean significativas para los estudiantes, ante lo cual también es necesario que el profesor tenga conocimiento en relación a los datos, así como del contexto en que estos se presentan.

El factor cultura adquiere protagonismo. Ambos, profesor y estudiante, están inmersos en un entorno sociocultural que forman parte de su identidad. El profesor no es un sujeto ajeno en el proceso de formación de sus estudiantes, su propia experiencia se hace presente al momento de guiarlos hacia la valoración y el análisis crítico de información en donde el contexto actúa como un factor mediador en el estudio estadístico no solo para el estudiante, sino también para el profesor, quien también debe continuar en constante formación.

Agradecimientos

Este trabajo fue realizado con el apoyo de la Beca de Manutención 2021 de la Pontificia Universidad Católica de Valparaíso.

Agradecimiento especial a la Dra. Elisabeth Ramos Rodríguez por sus revisiones y motivaciones para culminar este trabajo.

Referencias

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*, 99-125.

Aoyama, K., y Stephens, M. (2003). Graph interpretation aspects of statistical literacy: A Japanese perspective. *Mathematics Education Research Journal, 15*(3), 207-225. https://doi.org/10.1007/BF03217380

Arteaga, P., Batanero, C., Cañadas, G., y Contreras, M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números. Revista de Didáctica de las Matemáticas, 76,* 55-67.

Batanero, C., Díaz, C., Contreras, J. M., y Roa, R. (2013). El sentido estadístico y su desarrollo. *Números, 83,* 7-18.

Bertin, J. (1967). *Sémiologie graphique*. Gauthier-Villars.

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

Contreras, J., Molina-Portillo, E., Godino, J. D., y Batanero, C. (2017). Construcción de un cuestionario para evaluar la interpretación crítica de gráficos estadísticos por futuros profesores. En J. M. Muñoz, A. Arnal-Bailera, P. Beltrán-Pellicer, M. L. Callejo, y J. Carrillo (Eds.), *Investigaciones en Educación Matemática XXI* (207-216). Universidad de Zaragoza.

Curcio, F. R. (1987). Comprehension of Mathematical Relationships Expressed in Graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393. https://doi.org/10.5951/jresematheduc.18.5.0382

Curcio, F. R. (1989). *Developing Graph Comprehension. Elementary and Middle School Activities.* National Council of Teachers of Mathematics, Inc., 1906 Association Drive, Reston, VA 22091.

Díaz-Levicoy, D., Arteaga, P., y Batanero, C. (2015). Gráficos estadísticos y niveles de lectura propuestos en textos chilenos de educación primaria. En C. Fernández, M. Molina y N. Planas (Eds.), *Investigaciones en Educación Matemática XIX* (pp. 229-238). SEIEM.

Espinel, M. (2007). Construcción y Razonamiento de Gráficos Estadísticos en Formación de Profesores. *Investigación en Educación Matemática, 11*, 99-119.

Estrella, S., Olfos, R., y Mena-Lorca, A. (2015). El conocimiento pedagógico del contenido de estadística en profesores de primaria. *Educação e Pesquisa, 41*(2), 477-493. https://doi.org/10.1590/s1517-97022015041858.

Friel, S. N., Curcio, F. R., y Bright, G. W. (2001). Making Sense of Graphs. *Journal for Research in Mathematics Education, 32*(2), 124-158. https://doi.org/10.23077749671

Hernández, R., Fernández, C., y Baptista, P. (2014). *Metodología de la Investigación* (6.a edición). McGraw Hill.

Krippendorff, K. (2013). *Content analysis: an introduction to its methodology.* SAGE.

Mansilla, G., Soto, M., y Vivanco, C. (2020). *COVID-19: Implicancias y repercusiones en la seguridad hídrica.* Centro Regional del Agua para Zonas Áridas y Semiáridas de América Latina y el Caribe, CAZALAC.

Mautone, P. D., y Mayer, R. E. (2007). Cognitive Aids for Guiding Graph Comprehension. *Journal of Educational Psychology, 99*(3), 640-652. https://doi.org/10.1037/0022-0663.99.3.640

Ministerio de Educación de Chile. (2016). *Bases curriculares vigentes.* Unidad de Curriculum y Evaluación.

Pino, G. D. E. L., y Estrella, S. (2012). Educación estadística: relaciones con la matemática. *Revista de Investigación Educacional Latinoamericana, 49*(1), 53-64. http://dx.doi.org/10.7764/PEL.49.1.2012.5

Rodríguez, F., y Sandoval, P. (2012). Descodificación de tablas y gráficos estadísticos: Un estudio comparativo en profesores y alumnos de pedagogía en Enseñanza Básica. *Avaliação: Revista da Avaliação da Educação Superior, 17*(1), 207-235. https://doi.org/10.1590/S1414-40772012000100011

Shaughnessy, J., Garfield, J., y Greer, B. (1996). Data Handling. En A. Bishop, K. Clements, C. Keitel, J. Kilpatrick y C. Laborde (Eds.), *International Handbook of Mathematics Education. Kluwer International Handbooks of Education* (vol. 4, pp. 205-237). Springer. https://doi.org/10.1016/B978-0-12-375142-3.10006-9

Vásquez, C. (2021). Comprensión y Uso Docente de Gráficos Estadísticos por Futuros Profesores para Promover Competencias para la Sostenibilidad. *Paradigma, 41*(e1), 165-190. https://doi.org/10.37618/PARADIGMA.1011-2251.2021.p165-190.id1022

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

ARTÍCULOS DE INVESTIGACIÓN

https://doi.org/10.46219/rechiem.v13i4.77

RAZONAMIENTO ESTADÍSTICO EN EL CONTEXTO COVID-19: UNA PROPUESTA BASADA EN GEOGEBRA

STATISTICAL REASONING IN THE COVID-19 CONTEXT: A PROPOSAL WITH GEOGEBRA

Manuel González-Navarrete
magonzalez@ubiobio.cl
Universidad del Bío-Bío, Concepción, Chile

Iván Maldonado-Carrasco
ivan.maldonado06@inacapmail.cl
INACAP sede Puente Alto, Santiago, Chile

RESUMEN

En este trabajo presentamos una propuesta didáctica de análisis exploratorio de datos asociados a la evolución de la pandemia en las comunas de Chile. Las actividades son formuladas utilizando el software GeoGebra, en su versión clásica 6.0. Se seleccionaron 15 comunas de tres regiones para incluir un análisis univariado de los casos activos comunales, un análisis de regresión para identificar comunas cuyas evoluciones sean similares y análisis multivariado para comparar comunas en grupos. Finalmente, se ejemplifica un método de agrupamiento para clasificar comunas en función de sus casos activos por cada 100 mil habitantes.

PALABRAS CLAVE:
Razonamiento estadístico, Análisis exploratorio de datos, COVID-19, GeoGebra.

ABSTRACT

In this work we introduce a proposal of exploratory data analysis related to the evolution of the COVID-19 pandemic in the communes of Chile. The activities are formulated using the GeoGebra software, in its classic 6.0 version. We selected 15 communes from 3 regions to include a univariate analysis of active communal cases, a regression analysis to identify communes whose evolutions are similar, and multivariate analysis to compare communes in groups. Finally, an algorithm for community detection is used to classify communes based on their active cases per 100 thousand inhabitants.

KEYWORDS:
Statistical reasoning; Exploratory data analysis; COVID-19; GeoGebra.

Recibido: 26 de Mayo de 2021, Aceptado: 14 de Julio de 2021

1. Introducción

En el marco conceptual de la didáctica de la estadística, se destaca la jerarquía cognitiva de alfabetización estadística, razonamiento y pensamiento estadístico (Garfield, 2002). De acuerdo con Ben-Zvi y Garfield (2004), la alfabetización estadística incluye habilidades básicas de comprensión de la información estadística, destacando la capacidad de organizar datos, construir y presentar tablas y trabajar con diferentes representaciones de datos, agregando que el razonamiento estadístico se entiende como aquello que hacen las personas al razonar con ideas estadísticas y al dar sentido a la información estadística. El último nivel de pensamiento estadístico es más avanzado, puesto que implica una comprensión de por qué y cómo se realizan las investigaciones estadísticas. Esto incluye reconocer y comprender el proceso de investigación completo (Chance, 2002).

En esta misma línea, cobra relevancia el análisis exploratorio de datos (Tukey, 1977), entendido como la forma de resumir las características principales de los datos recolectados, utilizando a menudo gráficos estadísticos y otros métodos de visualización de datos. Así también se destaca el proceso de inferencia estadística informal, siendo definido como "la forma en que los estudiantes usan su conocimiento estadístico informal para crear argumentos que sustenten inferencias basados en los datos" (Zieffler et al., 2008, p. 44). La literatura en este sentido es extensa (ver, por ejemplo, a Arnold, 2008; Dvir y Ben-Zvi, 2021; Makar et al., 2011; Makar y Rubin, 2009, 2018; Pfannkuch, 2006; Rossman, 2008 y Wild et al., 2011).

De forma complementaria, la estadística ha jugado un rol esencial en el contexto de la actual pandemia de COVID-19. Diariamente se generan datos asociados a la evolución del virus en nuestro país, los cuales han sido dispuestos en una plataforma que incluye diversas variables a nivel comunal asociadas a la crisis sanitaria (Ministerio de Ciencia, 2021). El análisis de los datos disponibles es un desafío para la ciencia y una enorme oportunidad para la comunidad educativa en la enseñanza de la estadística, cobrando vital relevancia las palabras de Estrella (2017):

> *Nuestra actual sociedad requiere que los ciudadanos sean competentes en evaluar críticamente afirmaciones basadas en datos y en argumentar con fundamentos en la evidencia que entregan los datos. El estudio de la estadística provee a los estudiantes de herramientas, ideas y disposiciones para reaccionar inteligentemente a la ingente información del mundo que les rodea. (p. 173)*

De esta forma, se busca evidenciar la importancia del estudio de datos a nivel comunal de la evolución de la pandemia, permitiendo el desarrollo del razonamiento estadístico de forma contextualizada. De hecho, recientemente se definieron lineamientos y orientaciones específicas para la enseñanza y aprendizaje en tiempos de crisis (Propuestas Educación Mesa Social Covid-19, 2020), realizando una lectura de la priorización curricular elaborada el pasado año por la Unidad de Currículum y Evaluación del Ministerio de Educación (UCE MINEDUC). En particular para tercero medio, en el eje temático estadística y probabilidades, se sugiere una actividad basada en datos de coronavirus, argumentando que:

> *En el caso de Chile, se puede encontrar datos segmentados por región y comuna. Estos datos favorecen que los estudiantes puedan formular sus propias preguntas de interés y puedan analizar el comportamiento de los datos mediante los gráficos disponibles en distintas escalas. (Propuestas Educación Mesa Social Covid-19, 2020, p. 304)*

En este sentido, el currículum nacional chileno plantea cuatro unidades de trabajo en los niveles de tercero y cuarto medio en su plan electivo, estos se muestran en la Tabla 1. De esta forma, la presente propuesta se relaciona a las unidades 1 y 2.

Tabla 1 . Unidades y objetivos de aprendizaje (OA) del currículum nacional Nota. MINEDUC (2019, p. 151).

Unidades	**OA**
¿Qué dicen los gráficos? Análisis crítico de la información	Argumentar y comunicar decisiones a partir del análisis crítico de información presente en histogramas, polígonos de frecuencia, frecuencia acumulada, diagramas de cajón y nube de puntos, incluyendo el uso de herramientas digitales.
Media muestral, dispersión y correlación	Resolver problemas que involucren los conceptos de media muestral, desviación estándar, varianza, coeficiente de variación y correlación muestral entre dos variables, tanto de forma manuscrita como haciendo uso de herramientas tecnológicas digitales.
Situaciones o fenómenos que se modelan por medio de las distribuciones binomial y normal	Modelar fenómenos o situaciones cotidianas del ámbito científico y del ámbito social, que requieran el cálculo de probabilidades y la aplicación de las distribuciones binomial y normal.
Inferencia estadística	Argumentar inferencias acerca de parámetros (media y varianza) o características de una población, a partir de datos de una muestra aleatoria, bajo el supuesto de normalidad y aplicando procedimientos con base en intervalos de confianza o pruebas de hipótesis.

Por otra parte, la utilidad de las tecnologías y específicamente del software GeoGebra en la enseñanza de la estadística ha sido expuesta en diversos artículos (por ejemplo, Chance et al., 2007; Del Pino, 2013; García y Cuadros, 2013; Guncaga et al., 2019; Mellado y Marín, 2010; Prodromou, 2014; Zenteno Ruiz et al., 2020). Se suma a esto, el hecho de que GeoGebra en todas sus versiones es un software libre de código abierto y multiplataforma que está diseñado para todo nivel educativo escolar, y a su vez para educación superior tradicional como técnico-profesional. GeoGebra reúne dinámicamente geometría, álgebra, estadística y cálculo en registros gráficos, de análisis y de organización en hojas de cálculo (https://www.geogebra.org/). Además, existe una comunidad en la cual los usuarios de GeoGebra pueden interactuar a través de ella.

Por dichos motivos, el presente trabajo se basa en los datos de la evolución del virus para las comunas del país, resumidos en datos de casos activos disponibles en la plataforma covid19.ubiobio.cl. Particularmente, objetivamos exponer una propuesta didáctica utilizando las herramientas de GeoGebra clásico en su versión 6.0. El propósito es contribuir a la alfabetización y razonamiento estadístico basados en el análisis exploratorio de datos y el desarrollo de inferencia informal por parte de los estudiantes. La intención es realizar un estudio comparativo entre las comunas seleccionadas, por lo que sus medidas de resumen serán confrontadas, buscando similitudes y semejanzas en la evolución de la pandemia para las distintas comunas.

Para ello, en la Sección 2 seleccionamos cinco comunas de cada una de las siguientes regiones de Antofagasta (RA), Metropolitana (RM) y Los Ríos (RLR). Las actividades en la Sección 3 serán separadas en tres grupos: (a) análisis univariado, mediante medidas de tendencia central y dispersión, (b) análisis bivariado, mediante modelos de regresión lineal, y (c) análisis multivariado, con el objeto de comparar comunas dentro de la región o por periodos de tiempo específicos. Finalmente, basados en el análisis de regresión, proponemos una forma de agrupamiento utilizando el coeficiente de determinación entre los pares de comunas.

2. Recolección de datos comunales sobre la evolución del virus

Nuestra propuesta consiste en seleccionar un grupo de comunas para analizar la evolución de casos activos en cada una de ellas. Particularmente, utilizamos la serie de datos iniciada en abril de 2020 y terminando en el mismo mes de 2021. Los datos utilizados en esta propuesta pueden ser consultados en la plataforma de comparación comunal y regional (Universidad del Bío-Bío, 2021). La plataforma dispone de datos relacionados a casos activos y fallecidos a nivel comunal, además de las evoluciones de casos nuevos y fallecidos a nivel regional, permitiendo también descargar los datos seleccionados en el visor. Dado que buscamos comparar comunas, es importante conocer el tamaño de su población, puesto que influirá en la cantidad de casos que puedan reportarse diariamente, esto es, para una comuna más habitada se esperan más casos nuevos. En otras palabras, bajo las mismas condiciones de transmisión del virus, un grupo mayor de habitantes terminará con más personas contagiadas.

Las respectivas poblaciones de las comunas seleccionadas se resumen en la Tabla 2 y Figura 1, la cual incluye la media y mediana del número de habitantes para dichas comunas.

Tabla 2. Número de habitantes para las 15 comunas seleccionadas

Nota. Datos INE, proyección 2021 (Biblioteca del Congreso Nacional, 2021).

RA	**Antofagasta**	**Mejillones**	**San Pedro de Atacama**	**Sierra Gorda**	**Tocopilla**
Habitantes	425.725	14.776	10.434	1746	28.079
RM	La Pintana	Maipú	Puente Alto	Santiago	Vitacura
Habitantes	189.335	578.605	645.909	503.147	96.774
RLR	Futrono	Lanco	Los Lagos	Panguipulli	Valdivia
Habitantes	15.261	17.652	20.518	35.991	176.774

Nótese que la media está en torno a los 180 mil habitantes. Sin embargo, la mediana es cercana a los 36 mil habitantes, lo que corresponde específicamente a la población de Panguipulli. Otra observación que se desprende de esto es que la mayoría de las comunas en estudio poseen menos de 40 mil habitantes. Además, todas las comunas de la RM están por sobre la mediana de habitantes, siendo Puente Alto la que tiene la mayor población. Por otro lado, todas las comunas de la RLR están por debajo de la media de habitantes.

En este sentido, a la hora de comparar la evolución en diferentes comunas, es importante definir un indicador ponderado por la población respectiva. De esta forma, los datos utilizados serán los casos activos por cada 100 mil habitantes.

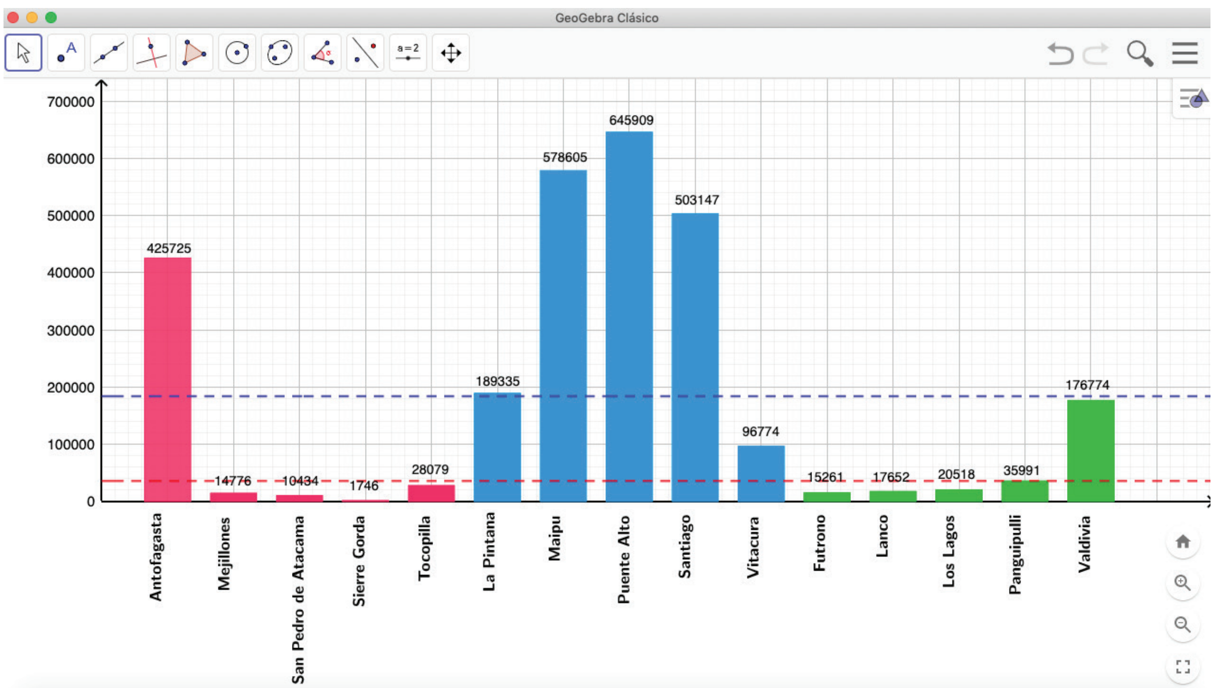


Figura 1. Gráfico de barras del número de habitantes para las 15 comunas seleccionadas

Nota. Elaboración propia. Las líneas horizontales representan la media (azul) y mediana (roja) del número de habitantes para las 15 comunas.

2.1 Visualización de datos históricos en covid19.ubiobio.cl

La plataforma utilizada es de fácil manipulación y sus diversas herramientas favorecen la visualización y adquisición de los datos, los cuales son constantemente actualizados desde la base de datos del Ministerio de Ciencia.

En la Figura 2 vemos la pantalla inicial de la plataforma, en ella se pueden seleccionar comunas escribiendo sus nombres, o bien, escoger una región para comparar todas sus comunas. Las gráficas generadas pueden ser visualizadas por periodos específicos a criterio del usuario. Al descargar en formato CSV, el usuario dispone de los datos de población, casos activos y casos activos por cada 100 mil habitantes para las comunas seleccionadas.

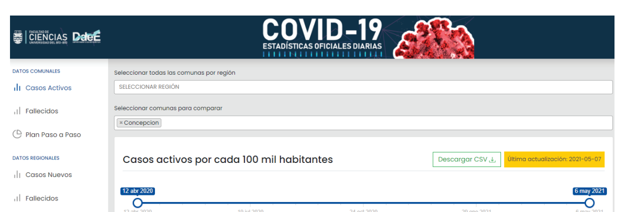


Figura 2 . Pantalla de inicio de la plataforma de datos comunales de la Universidad del Bío-Bío.

Nota. Obtenido de covid19.ubiobio.cl.

Utilizando la plataforma, en la Figura 3 se presentan las cinco comunas de la RM seleccionadas para este estudio: La Pintana, Maipú, Puente Alto, Santiago y Vitacura. Es posible notar que estas tienen una evolución similar para el indicador seleccionado. Se observa un momento crítico (*peak*) en torno a los meses de mayo y junio de 2020, con épocas de primavera y verano con valores bajos.

Se evidencia que La Pintana ha presentado los índices más altos y actualmente (hasta el día 6 de mayo) junto a Puente Alto presentan la situación más delicada de estas cinco comunas. En cambio, Vitacura se ha mostrado con una situación más controlada.

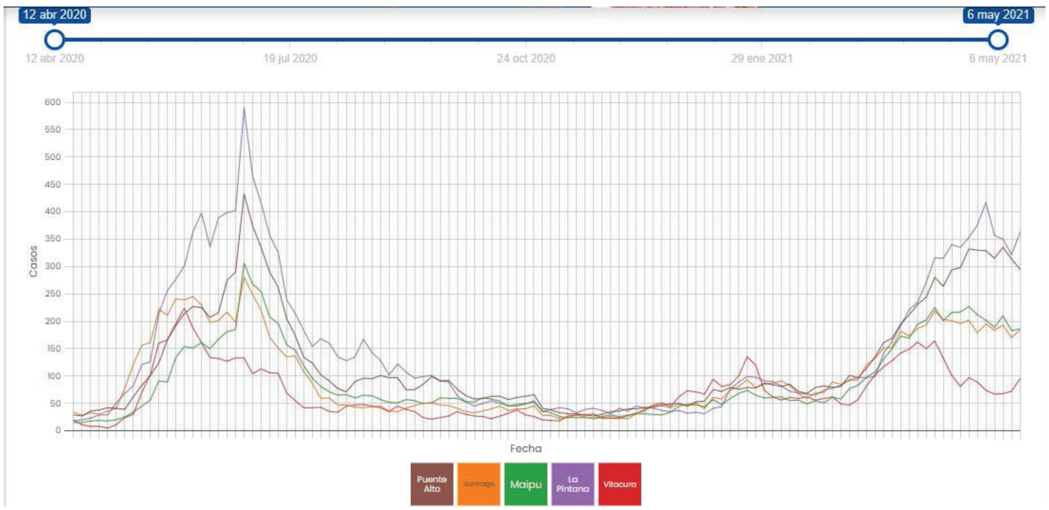


Figura 3 . Evolución de casos activos por cada 100 mil habitantes para las comunas de la RM

Nota. Generado por covid19.ubiobio.cl.

En la Figura 4, al considerar las comunas de la RLR, Futrono, Lanco, Los Lagos, Panguipulli y Valdivia, se observa un momento crítico para la estación de verano 2021, antecedido por una primavera con clara tendencia al alza. Mencionamos que la comuna de Futrono presenta el valor histórico más alto de las cinco comunas, el que fue alcanzado los primeros días del mes de marzo. Se destaca también la comuna de Los Lagos con una tendencia al alza constante durante los últimos meses, finalizando una larga cuarentena en el mes de abril de 2021, pero siendo actualmente la comuna con los peores valores.

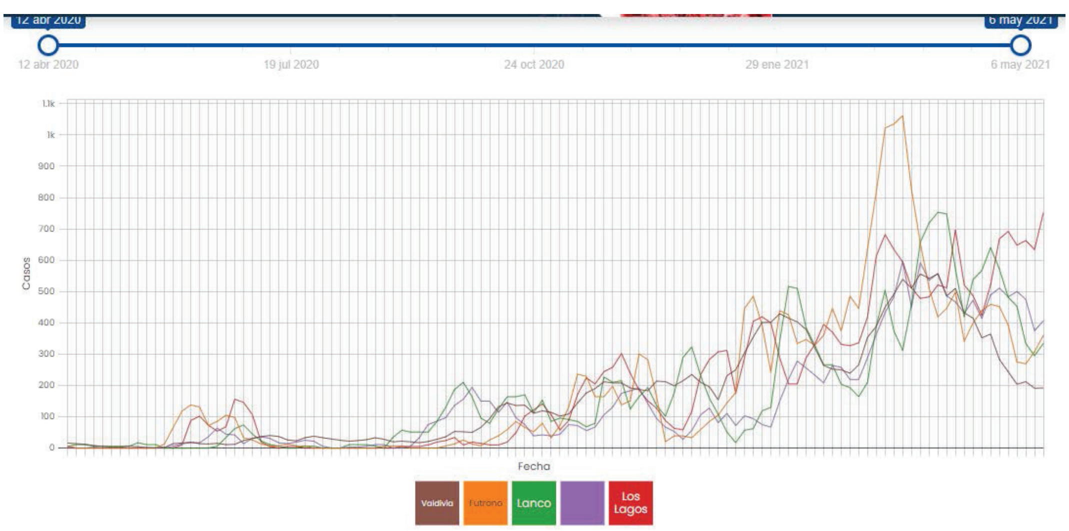


Figura 4 . Evolución de casos activos por cada 100 mil habitantes para las comunas de la RLR

Nota. Generado por covid19.ubiobio.cl.

Finalmente, al observar las evoluciones para las comunas de Antofagasta, Mejillones, San Pedro de Atacama, Sierra Gorda y Tocopilla, podemos encontrar diferencias notorias. Particularmente, la comuna de Sierra Gorda presenta un comportamiento disonante con las otras comunas, esto puede deberse al tamaño de su población (ver Tabla 2). En el caso de las cuatro restantes, tuvieron un periodo crítico en los meses de enero y febrero de 2021.

El caso de Mejillones representa el *peak* más alto de todas las comunas, en torno a los 900 casos activos por cada 100 mil habitantes a comienzos del mes de febrero de 2021.

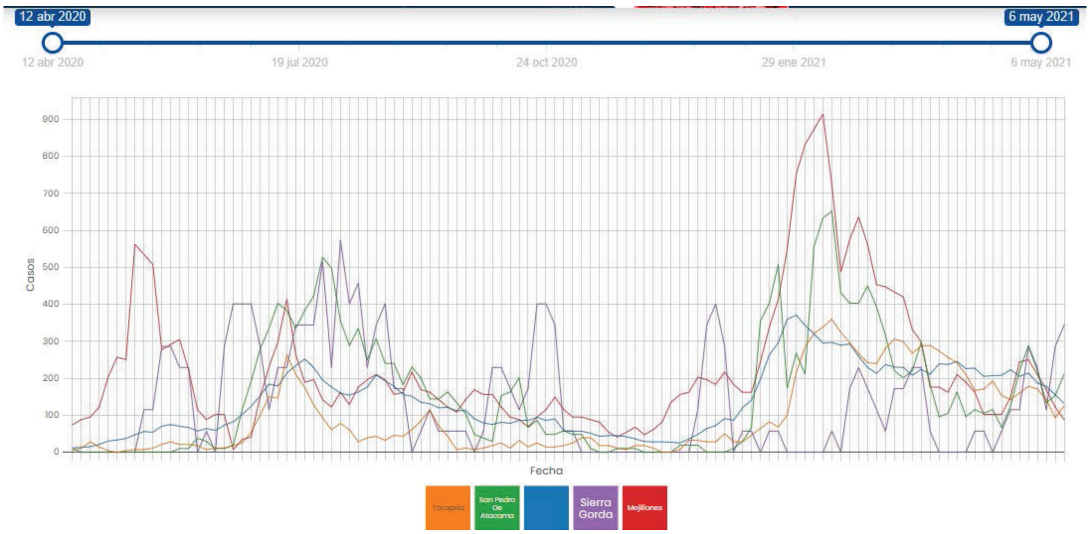


Figura 5 . Evolución de casos activos por cada 100 mil habitantes para las comunas de la RA

Nota. Generado por covid19.ubiobio.cl.

3. Análisis exploratorio con GeoGebra

Una vez seleccionadas las 15 comunas, procedemos a descargar los datos en formato CSV. En este punto, se debe confeccionar una planilla simplificada para procesar en GeoGebra, dicha tarea puede ser desarrollada en Excel.

Las herramientas de GeoGebra que serán utilizadas son: (1) Análisis de una variable, (2) Análisis de regresión de dos variables y (3) Análisis multivariable (ver Figuras 21, 26 y 29, respectivamente). Los detalles de la manipulación requerida para las actividades pueden ser consultados en el Apéndice. En esta sección nos concentraremos en exponer el tipo de resultados y conclusiones a los que es posible llegar, indicando por tanto las posibilidades de desarrollar alfabetización y razonamiento estadístico.

3.1 Caracterización de variables unidimensionales

A modo de ejemplo, consideraremos las comunas que visualmente presentan *peak* más altos en sus respectivas regiones, esto es Mejillones (RA), La Pintana (RM) y Futrono (RLR). La serie de datos utilizada corresponde al periodo del 13-04-20 al 23-04-21.

En primer lugar, visualizamos el histograma, y a continuación el diagrama de cajas con su gráfico Q-Q para la distribución normal. Los detalles de cada construcción y sus propiedades pueden ser consultados en Gibbons y Chakraborti (2003), Triola (2009) y Tukey (1977).

Con el histograma de la Figura 6 podemos evidenciar que la mayor parte del tiempo la comuna de Mejillones ha tenido índices de casos activos por cada 100 mil habitantes en el intervalo de 150 a 180, esto se puede visualizar también en la tabla de frecuencias. Observando la distribución de esta comuna, vemos asimetría hacia el lado izquierdo (positiva). Desde las estadísticas se observa que la media (236.13) es mayor a la mediana (162), en esta misma línea es posible identificar las medidas de posición: cuartil 1 (108), cuartil 2 (162) y cuartil 3 (291).

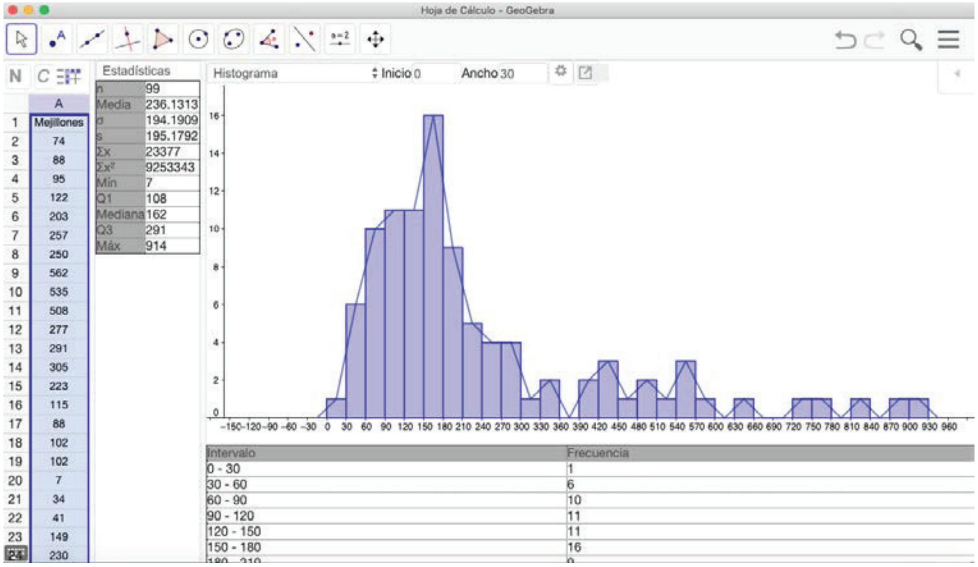


Figura 6. Histograma para los casos activos por cada 100 mil habitantes en la comuna de Mejillones

Nota. Elaboración propia.

Las estadísticas mencionadas pueden ser visualizadas en el gráfico de cajas (*boxplot*) de la Figura 7, observando la existencia de datos atípicos (*outliers*). Bajo el *boxplot* observamos el gráfico Q-Q (cuantil-cuantil), el cual nos compara dos distribuciones: una empírica (comuna de Mejillones) versus la distribución teórica (normal o de Gauss); en este caso los puntos están representando los datos para la comuna de Mejillones y la recta indica normalidad. Nótese que el histograma y el gráfico Q-Q señalan que la distribución empírica no aparenta ser una distribución normal.

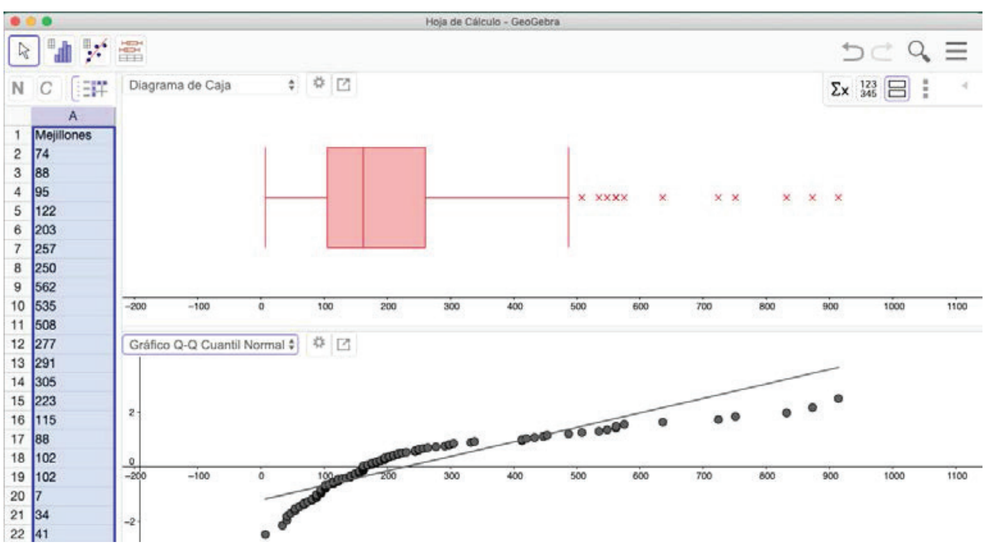


Figura 7 . Diagrama de caja (boxplot) y gráfico Q-Q para los casos activos por cada 100 mil habitantes en la comuna de Mejillones

Nota. Elaboración propia.

A continuación, visualizamos el histograma, diagrama de cajas y gráfico Q-Q para la comuna de La Pintana. Considerando el histograma de la Figura 8, es posible observar que la mayor parte del tiempo la comuna de La Pintana ha tenido índices de casos activos en el intervalo de 30 a 60, esta distribución presenta cierta asimetría positiva. Al mismo tiempo, desde el resumen de estadísticas se puede distinguir que la media (142.19) es mayor a la mediana (92), es posible identificar por consiguiente las medidas de posición: cuartil 1 (45), cuartil 2 (92) y cuartil 3 (203).

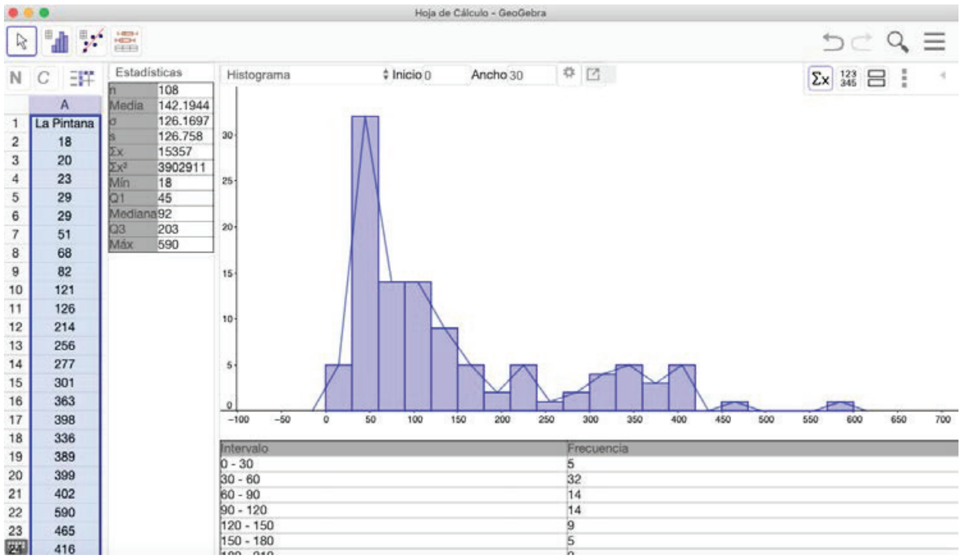


Figura 9 . Diagrama de caja (boxplot) y gráfico Q-Q para los casos activos por cada 100 mil habitantes en la comuna de La Pintana

Nota. Elaboración propia.

A continuación, visualizamos el histograma, diagrama de cajas y gráfico Q-Q para Futrono. En la Figura 10 se evidencia que la mayor parte del tiempo Futrono ha pasado con bajos índices de casos activos, esto debido a que el intervalo de 0 a 30 se presenta con mayor frecuencia. Además, la media (188.57) es mayor a la mediana (75.5) y las medidas de posición: cuartil 1 (7), cuartil 2 (75.5) y cuartil 3 (344).

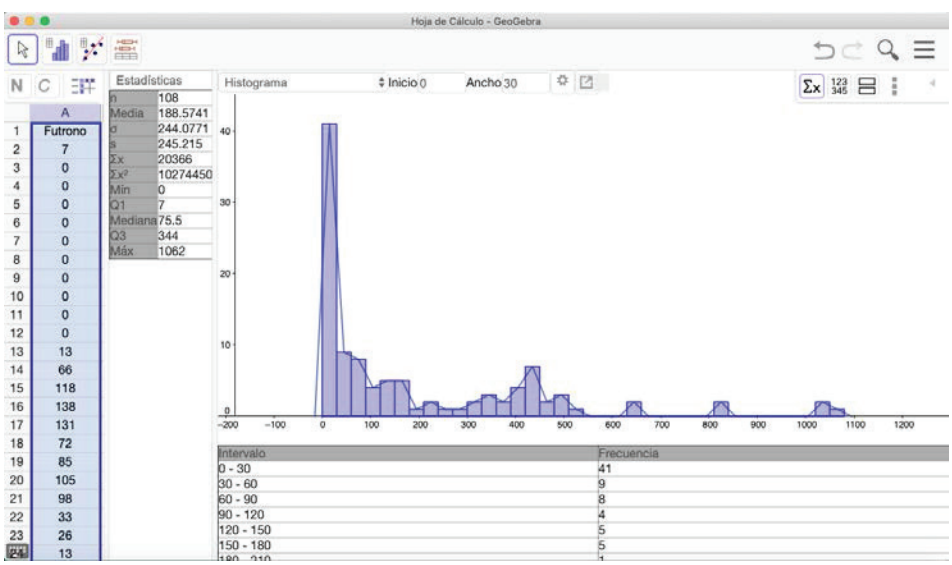


Figura 10. Histograma para los casos activos por cada 100 mil habitantes en la comuna de Futrono

Nota. Elaboración propia.

Observando el gráfico de cajas (*boxplot*) de la comuna de Futrono, visualizamos la existencia de tres datos atípicos (*outliers*). Bajo el *boxplot* observamos el gráfico Q-Q (cuantil-cuantil), con lo que se concluye con mayor seguridad que los datos no aparentan una distribución normal.

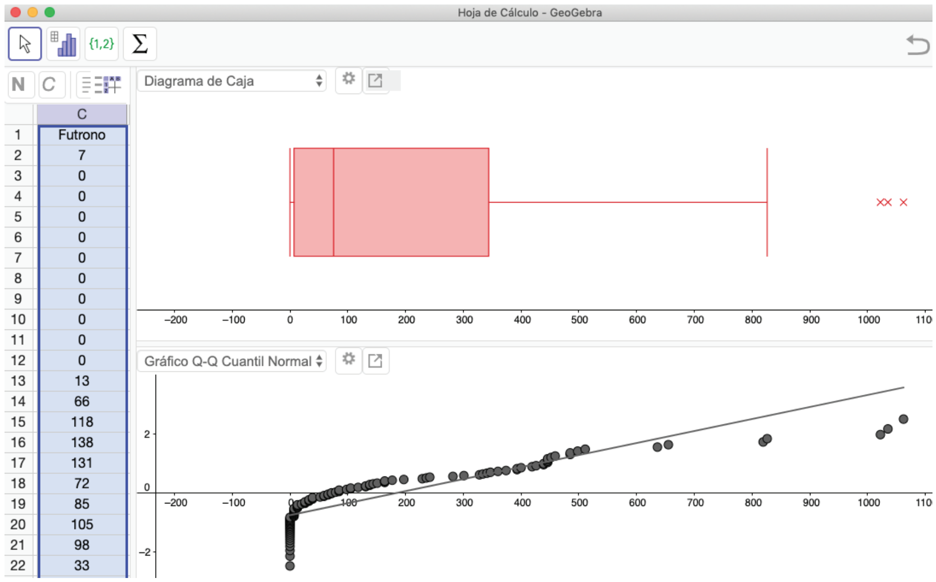


Figura 11 . Diagrama de caja (boxplot) y gráfico Q-Q para los casos activos por cada 100 mil habitantes en la comuna de Futrono
Nota. Elaboración propia.

3.2 Análisis de regresión

Considerando la RM, realizaremos un análisis de regresión para identificar qué tan similares (o disímiles) han sido las evoluciones del indicador en estudio para las cinco comunas seleccionadas.

Comenzando con Puente Alto (eje X) versus La Pintana (eje Y), en la Figura 12 vemos el resumen entregado por GeoGebra, en el que se presenta la ecuación de la recta que describe la regresión lineal, esto es:

$$y=1.3146x-14.4293.$$

Se entiende que la recta tiene pendiente positiva, lo que indica una correlación directa entre las comunas. En el caso del intercepto, este es cercano a -14. Además, el programa nos entrega el valor de R^2 (0.9309), llamado coeficiente de determinación, el cual, mientras más cercano a 1 sea, indica mayor poder predictivo para la recta de regresión. En GeoGebra es posible obtener otros tipos de regresión (ver Figura 28), pero nos enfocamos en el caso lineal para comenzar. El coeficiente de determinación será utilizado en la Sección 3.3.1, para construir un indicador de dependencia geográfica de la evolución del virus.

En este sentido, es posible utilizar la técnica de regresión lineal para analizar tendencias en los datos. De esta forma, basta cruzar los datos de una comuna con una secuencia de fechas o secuencia numérica consecutiva, que represente los días del periodo analizado. La idea es identificar una recta de tendencia que podría indicar un alza, estabilidad o disminución de los casos activos en el periodo analizado. Este tipo de estudios es de utilidad si se quiere analizar la época presente, por las últimas dos semanas, a modo de proyectar la evolución para los próximos días.

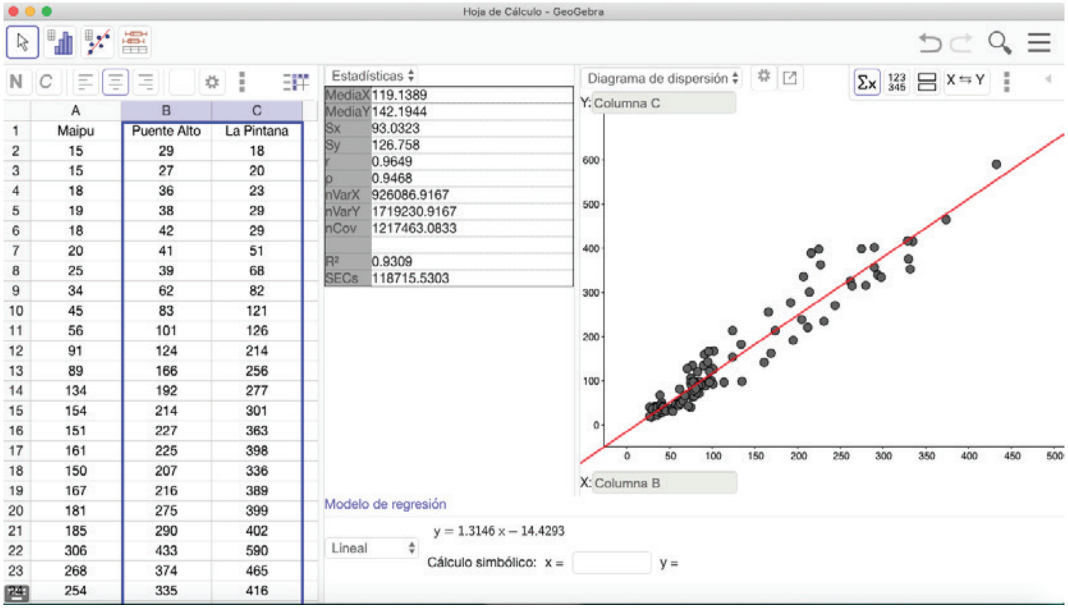


Figura 12 . Análisis de regresión para los casos activos por cada 100 mil habitantes en las comunas de Puente Alto y La Pintana

Nota. Elaboración propia.

Replicando este análisis con todos los pares de comunas de la RM, obtenemos un resumen de los valores de R^2 en la Tabla 3. Se observa que Vitacura es la comuna con un comportamiento diferente a las otras cuatro comunas. Las comunas con mayor índice de dependencia son Puente Alto y La Pintana. Finalmente, los gráficos de dispersión y la recta regresora están incluidos en la Figura 13.

Tabla 3. Valores de R^2 obtenido en GeoGebra mediante análisis de regresión lineal

Nota. Elaboración propia.

	La Pintana	**Maipú**	**Puente Alto**	**Santiago**	**Vitacura**
La Pintana	1	0.8794	0.9683	0.7834	0.5056
Maipú	0.8794	1	0.9309	0.8094	0.4954
Puente Alto	0.9683	0.9309	1	0.7976	0.4776
Santiago	0.7834	0.8094	0.7976	1	0.7547
Vitacura	0.5056	0.4954	0.4776	0.7547	1

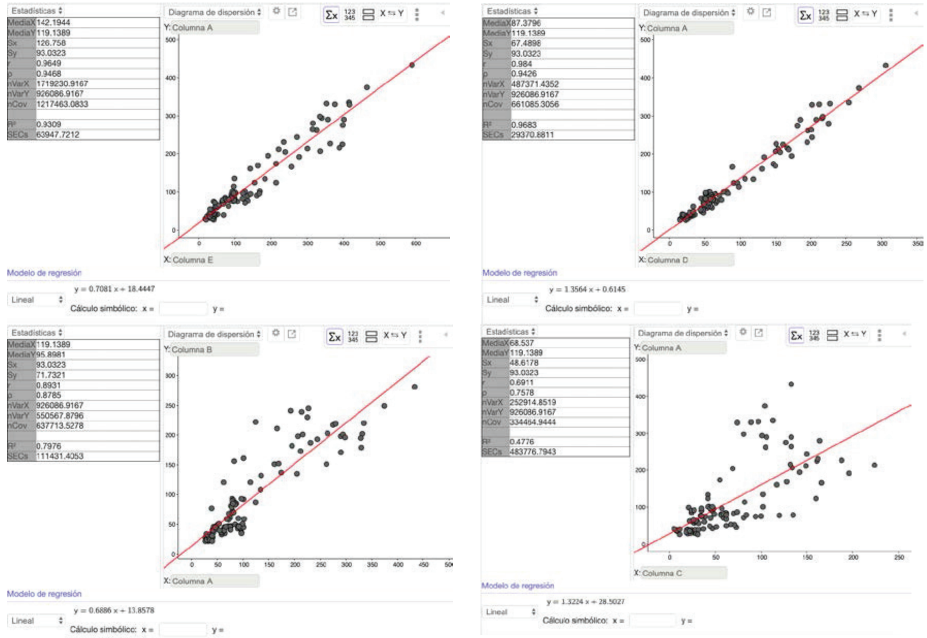


Figura 13. Análisis de regresión para los casos activos por cada 100 mil habitantes en las comunas de Puente Alto con La Pintana, Maipú, Vitacura y Santiago. Nota. Elaboración propia. . El orden es de izquierda a derecha y de arriba hacia abajo.

3.3 Análisis multivariado y agrupamiento

En esta línea, podemos hacer una comparación de las comunas en virtud de los diagramas de cajas y bigotes (*boxplots*) asociados. Proponemos dos formas de agrupar las comunas para ser comparadas. En primer lugar, analizamos por regiones, cuyos detalles están en las Figuras 14 y 15. Posteriormente, comparamos las tres comunas estudiadas en la Sección 3.1 (ver Figura 16). Finalmente, utilizando el coeficiente de determinación R^2 obtenido en la Sección 3.2, proponemos un método de agrupamiento para clasificar las evoluciones de las 15 comunas seleccionadas.

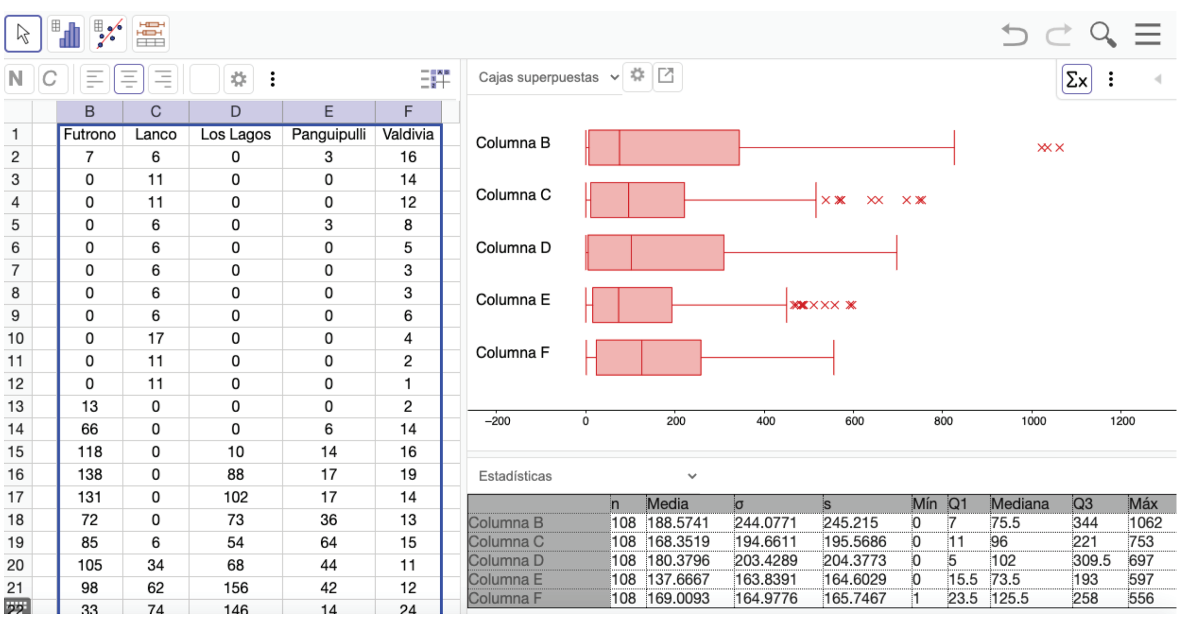


Figura 14 . Boxplot para los casos activos por cada 100 mil habitantes en las comunas de la RLR Nota. Elaboración propia.

En la Figura 14, nótese que la comuna de Futrono presenta una caja más grande y valores atípicos que se escapan bastante en comparación a las otras comunas. Por otro lado, todas las comunas presentan un primer cuartil cercano a cero.

De forma análoga generamos los diagramas de caja para los casos activos por cada 100 mil habitantes en las comunas de las RM y RA, los detalles están en la Figura 15, cuya identificación se encuentra en la Tabla 4.

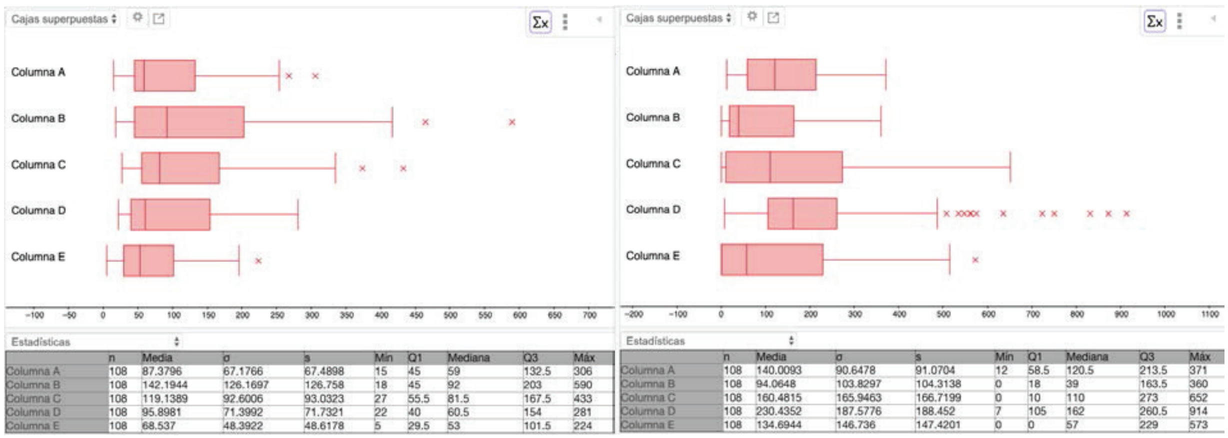


Figura 15. Boxplot para las comunas de la RM (izquierda) y la RA (derecha)

Nota. Elaboración propia.

Tabla 4. Codificación usada en la Figura 15, comunas de la RM y RA. Nota. Elaboración propia.

Columna	**Comunas RM**	**Comunas RA**
Columna A	Maipú	Antofagasta
Columna B	La Pintana	Tocopilla
Columna C	Puente Alto	San Pedro de Atacama
Columna D	Santiago	Mejillones
Columna E	Vitacura	Sierra Gorda

Siguiendo con el análisis multivariado, es interesante estudiar en comparativa las tres comunas analizadas en la Sección 3.1. Particularmente, consideraremos la estación de verano, desde el 21-12-20 al 22-03-21, correspondientes a 27 registros. Mostramos un resumen de estadísticas en la Tabla 5 y los respectivos *boxplot* en la Figura 16.

Al comparar media, mediana y cuartiles, es posible observar que la comuna de La Pintana durante el periodo de verano fue menos afectada que las comunas de Mejillones y Futrono. En esta línea, es posible calcular el coeficiente de variación de estas comunas en el periodo de verano. Con los resultados de la Tabla 5, es posible observar que La Pintana presenta una mayor homogeneidad de los datos con respecto a las otras dos comunas y la comuna con mayor heterogeneidad sería Futrono.

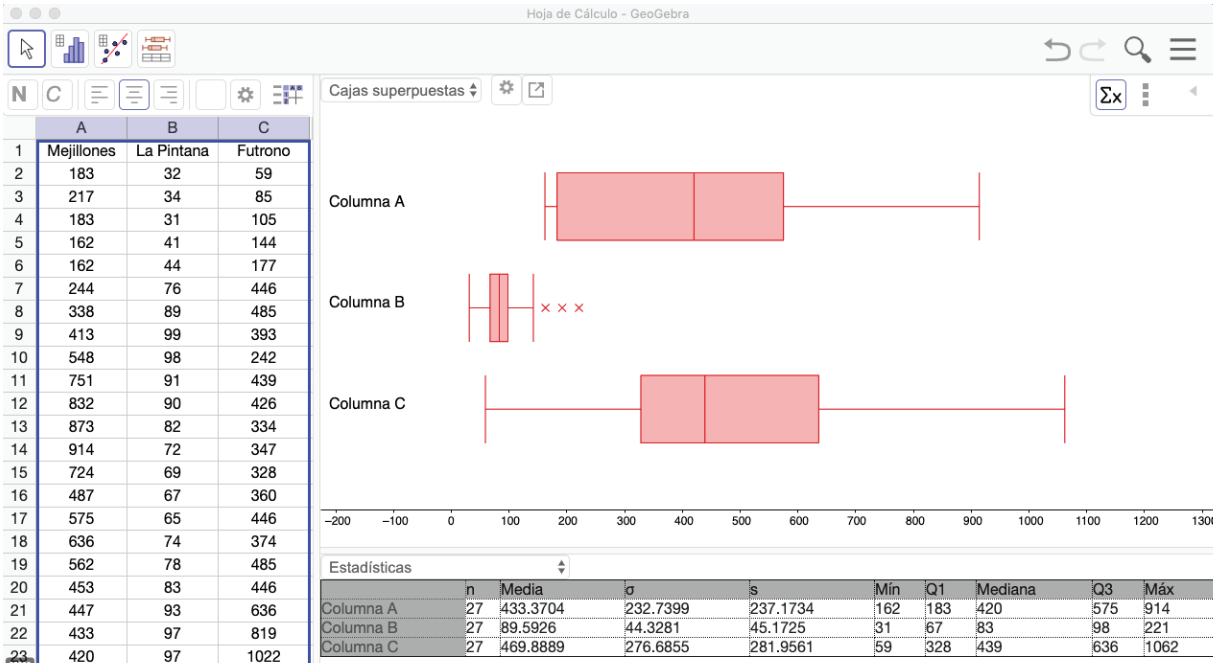


Figura 16. Boxplot para los casos activos por cada 100 mil habitantes en las comunas de Mejillones, La Pintana y Futrono, para la temporada de verano Nota. Elaboración propia.

Tabla 5. Coeficiente de variación en periodo de verano de las comunas de Mejillones, La Pintana y Futrono para los casos activos por cada 100 mil habitantes. Nota. Elaboración propia.

	Desviación estándar	**Media**	**Coeficiente de variación**
Mejillones	237.17	433.37	0.547
La Pintana	45.17	89.59	0.504
Futrono	281.96	469.89	0.601

3.3.1 Agrupamiento basado en el coeficiente de determinación de la regresión lineal

En esta sección se incluye un análisis de agrupamiento con respecto a todas las comunas seleccionadas. Esto será basado en el coeficiente de determinación R^2 obtenido del análisis de regresión, como fue explicado en la Sección 3.2.

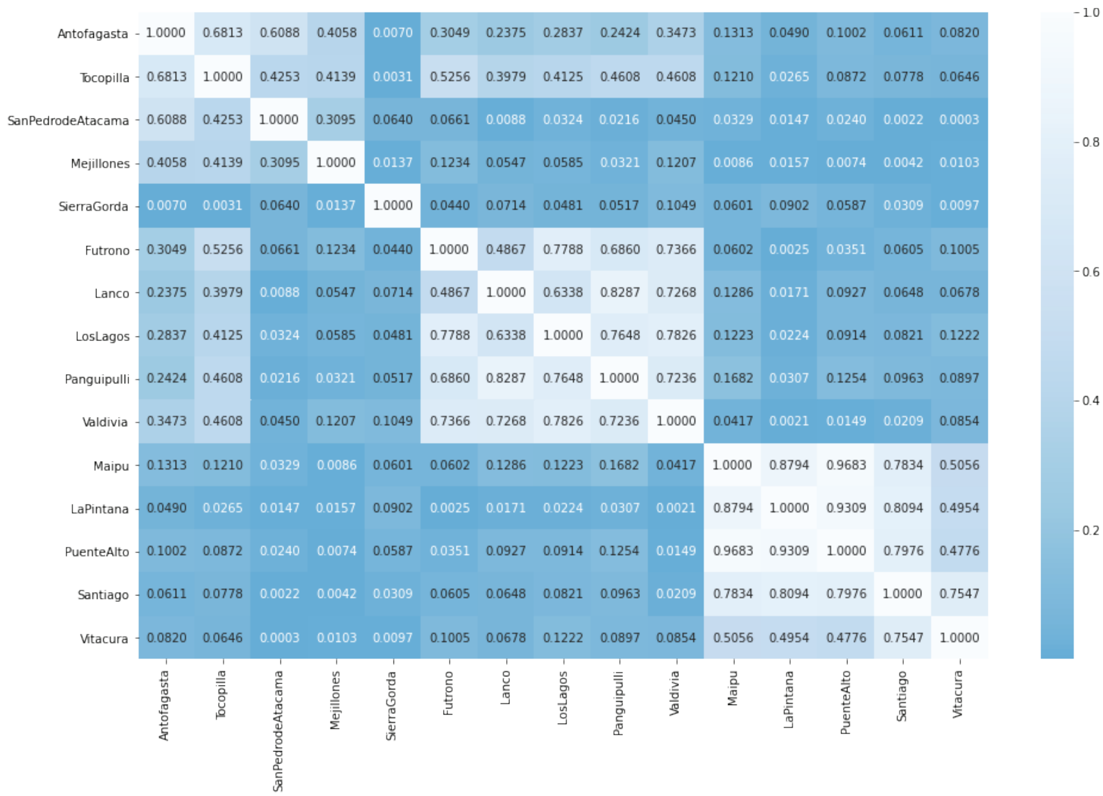


Figura 17 . Resumen de los valores de R^2 ᵛasociados a los análisis de regresión

Nota. Elaboración propia (usando Python).

Las comunas de la RM presentan indicadores de dependencia bastante altos y se podrían agrupar en un clúster, así como también se podrían agrupar las comunas de la RLR. Sin embargo, las comunas de la RA presentan menores índices de dependencia; incluso, la comuna de Sierra Gorda parece no estar significativamente correlacionada con las demás comunas de su región. Como hemos mencionado, la población de dicha comuna influye en su evolución y en su forma diferenciada de las demás.

Mencionamos una posible herramienta para obtener una clasificación en clúster, el llamado algoritmo de Louvain, propuesto por Blondel et al. (2008). Para utilizar el método, se necesita una ponderación entre elementos (comunas), que podría ser considerada con el coeficiente de determinación de la regresión lineal.

De esta forma, un par de comunas con coeficiente más alto tendrán más probabilidad de ser identificadas en un mismo clúster. Esto ocurre con las comunas de la RM, por ejemplo. De hecho, al aplicarlo en busca de clústeres para las 15 comunas, se obtienen los cuatro clústeres identificados en la Tabla 6.

Tabla 6. Clasificación de las 15 comunas mediante el algoritmo de Louvain

Nota. Elaboración propia, usando el software R.

RA	Antofagasta	Mejillones	San Pedro de Atacama	Tocopilla	Sierra Gorda
Clúster	1	1	1	1	2
RM	La Pintana	Maipú	Puente Alto	Santiago	Vitacura
Clúster	3	3	3	3	3
RLR	Futrono	Lanco	Los Lagos	Panguipulli	Valdivia
Clúster	4	4	4	4	4

4. Comentarios finales

En el presente artículo realizamos una propuesta de análisis exploratorio de las series de evolución de casos activos para 15 comunas de Chile, entre abril de 2020 y abril de 2021. En este sentido, los análisis pueden ser realizados considerando periodos diferentes de tiempo, como el efectuado para la estación de verano en la Figura 16.

Para el caso de la RM, es importante complementar los análisis con la discusión del artículo de Mena et al. (2021), el cual caracteriza la influencia de la variable socioeconómica en la tasa de mortalidad para las comunas de dicha región.

Otras actividades pueden ser basadas en los datos regionales disponibles en la plataforma covid19.ubiobio.cl. En este caso, es posible realizar comparaciones entre sintomáticos y asintomáticos. Además, dichos datos poseen una frecuencia diaria, por lo que se pueden aplicar técnicas de suavizamiento, como las medias móviles. Sumado a esto, en el repositorio del Ministerio de Ciencia se encuentran disponibles los datos de positividad y el avance del proceso de vacunación para las regiones y comunas del país.

Complementado a lo que hemos indicado, esta propuesta es aplicable en tercero y cuarto medio de Educación Media en el contexto chileno, considerando los planes electivos, tanto por la pertinencia de las herramientas GeoGebra y Excel, como por la contingencia de los datos COVID-19. Por este mismo motivo, es totalmente factible aplicar esta propuesta en asignaturas de educación superior. En este sentido, las diversas actividades pueden ser adaptadas a los contextos en función de la disponibilidad de datos.

Como trabajo futuro, pretendemos implementar esta propuesta en el aula escolar, para comprobar los beneficios o desventajas que puedan presentarse. De esta forma, el estudio de clases se vuelve una interesante línea de investigación para ser explorada (Estrella y Estrella, 2020; Isoda y Olfos, 2020; Olfos et al., 2015).

Agradecimientos

Los autores agradecen a los revisores anónimos por sus valiosos aportes al mejoramiento de este trabajo y a Nelly Gómez (UBB) por sus comentarios y correcciones. Este trabajo fue parcialmente financiado por Fondecyt Iniciación nº 11200500.

Referencias

Arnold, P. (2008). Developing new statistical content knowledge with secondary school mathematics teachers. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference* (pp. 1-6). International Commission on Mathematical Instruction and International Association for Statistical Education.

Ben-Zvi, D., y Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. En *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3-15). Springer. https://doi.org/10.1007/1-4020-2278-6

Biblioteca del Congreso Nacional. (2021, 20 de mayo). *Reportes comunales.* https://www.bcn.cl/siit/reportescomunales/index.html

Blondel, V., Guillaume, J., Lambiotte, R., y Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment,* (10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3), 1-14. https://doi.org/10.1080/10691898.2002.11910677

Chance, B., Ben-Zvi, D., Garfield, J., y Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal, 1*(1), 1-27. https://doi.org/10.5070/T511000026

Del Pino, J. (2013). El uso de GeoGebra como herramienta para el aprendizaje de las medidas de dispersión. *Probabilidad Condicionada: Revista de didáctica de la estadística, 2*, 243-250.

Dvir, M., y Ben-Zvi, D. (2021). Informal statistical models and modeling. *Mathematical Thinking and Learning*, 1-21. https://doi.org/10.1080/10986065.2021.1925842

Estrella, S. (2017). Enseñar estadística para alfabetizar estadísticamente y desarrollar el razonamiento estadístico. En A. Salcedo (Ed.), *Alternativas Pedagógicas para la Educación Matemática del Siglo XXI* (pp. 173-194). Centro de Investigaciones Educativas, Escuela de Educación. Universidad Central de Venezuela. http://saber.ucv.ve/bitstream/123456789/15712/1/Alternativas%20Pedagogicas%20para%20la%20Educaci%C3%B3n%20Matematica%20S%20XXI.pdf

Estrella, S., y Estrella, P. (2020). Representaciones de datos en estadística: de listas a tablas. *Revista Chilena de Educación Matemática, 12*(1), 21-34. https://doi.org/10.46219/rechiem.v12i1.20

García, G., y Cuadros, P. (2013). Estrategias para mejorar la enseñanza de la Estadística con GeoGebra. En Sociedad de Educación Matemática Uruguaya (Ed.), *VII Congreso Iberoamericano de Educación Matemática* (pp. 6335-6342). Autor.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3), 1-12. https://doi.org/10.1080/10691898.2002.11910676

Gibbons, J. D., y Chakraborti, S. (2003). *Nonparametric Statistical Inference.* Marcel Dekker.

Guncaga, J., Zawadowski, W., y Prodromou, T. (2019). Visualisation of Selected Mathematics Concepts with Computers – the Case of Torricelli's Method and Statistics. *European Journal of Contemporary Education, 8*(1), 69-91. https://doi.org/10.13187/ejced.2019.1.69

Isoda, M., y Olfos, R. (2020). *Teaching multiplication with Lesson Study: Japanese and Ibero-American Theories for Mathematics Education.* Springer. https://doi.org/10.1007/978-3-030-28561-6

Makar, K., Bakker, A., y Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*, 152-173. https://doi.org/10.1080/10986065.2011.538301

Makar, K., y Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105.

Makar, K., y Rubin, A. (2018). Learning about statistical inference. En D. Ben-Zvi, K. Makar y J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261-294). Springer. https://doi.org/10.1007/978-3-319-66195-7_8

Mellado, A., y Marín, L. A. (2010). Distribuciones estadísticas: Un ejemplo de uso de GeoGebra en enseñanza universitaria. *Épsilon, 74*, 33-42.

Mena, G. E., Martínez, P. P., Mahmud, A. S., Marquet, P. A., Buckee, C. O., y Santillana, M. (2021). Socioeconomic status determines covid-19 incidence and related mortality in Santiago, Chile. *Science*, 372(6545). https://doi.org/10.1126/science.abg5298

Ministerio de Ciencia. (2021, 20 de mayo). *Datos-COVID19* [GitHub dataset]. https://github.com/MinCiencia/Datos-COVID19

Ministerio de Educación de Chile. (2019). Bases Curriculares 3° y 4° año medio. Autor. https://www.curriculumnacional.cl/614/articles-89597_recurso_10.pdf

Olfos, R., Estrella, S., y Morales, S. (2015). Clase pública de un estudio de clases de estadística: Una instancia de cambio de creencias en los profesores. *Revista Electrónica Educare, 19*(3), 1-17. https://doi.org/10.15359/ree.19-3.21

Pfannkuch, M. (2006). Informal inferential reasoning. En A. Rossman y B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics* (pp. 1-6). International Association for Statistics Education.

Prodromou, T. (2014). GeoGebra in Teaching and Learning Introductory Statistics. *Electronic Journal of Mathematics and Technology, 8*(5), 363-376.

Propuestas Educación Mesa Social Covid-19. (2020). *Didácticas para la proximidad: aprendiendo en tiempos de crisis*. Pontificia Universidad Católica de Chile y Universidad de Chile. https://www.pucv.cl/uuaa/site/docs/20200723/20200723181922/dida__cticas_para_la_proximidad___aprendiendo_en_tiempos_de_crisis.pdf

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal, 7*(2), 5-19.

Triola, M. F. (2009). *Estadística* (10.a Ed.). Pearson Education.

Tukey, J. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Universidad del Bío-Bío. (2021, 20 de mayo). *Plataforma de comparación comunal datos covid-19.* http://covid19.ubiobio.cl/index.html

Wild, C. J., Pfannkuch, M., Regan, M., y Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*, 247-295. https://doi.org/10.1111/j.1467-985X.2010.00678.x

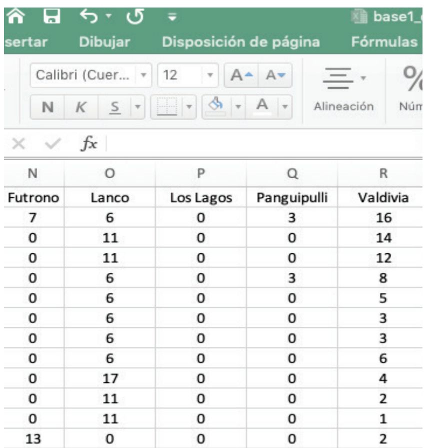
Zenteno Ruiz, F. A., Rivera Espinoza, T. A., y Pariona Cervantes, D. J. (2020). Tratamiento de las medidas de dispersión por medio del software GeoGebra. *Universidad y Sociedad, 12*(1), 244-250.

Zieffler, A., Garfield, J., Delmas, R., y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40-58.

Apéndice: manipulación en el ambiente de trabajo de GeoGebra

En esta sección se presentan las rutinas en GeoGebra utilizadas para desarrollar el análisis exploratorio de datos de la Sección 3. Para más detalles, visitar https://youtube.com/playlist?list=PL2eb1_IpIPd-PCwip0-5ijiw0PUUi21DM .

Se destaca que la información descargada en formato CSV se consolidó en una planilla de cálculo, obteniendo una tabla como en la Figura 18.



Futrono	Lanco	Los Lagos	Panguipulli	Valdivia
7	6	0	3	16
0	11	0	0	14
0	11	0	0	12
0	6	0	3	8
0	6	0	0	5
0	6	0	0	3
0	6	0	0	3
0	6	0	0	6
0	17	0	0	4
0	11	0	0	2
0	11	0	0	1
13	0	0	0	2

Figura 18. Encabezado de datos organizados en planilla de cálculo. Nota. Elaboración propia.

A continuación, se prepara el ambiente de trabajo de GeoGebra (Figura 19). Se deja habilitada sola la hoja de cálculo, dejando deshabilitada vista algebraica y gráfica (ver Figura 20).

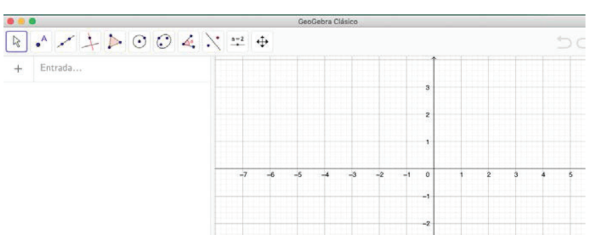


Figura 19. Vista por defecto en GeoGebra Clásico 6.0
Nota. Elaboración propia.

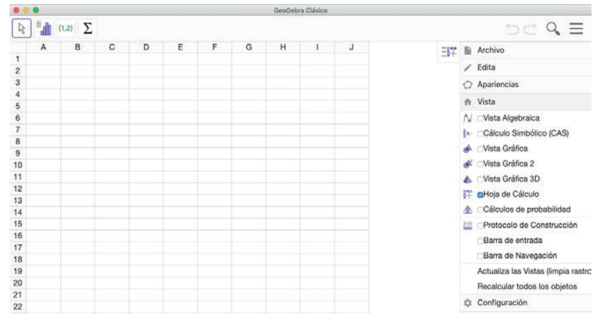
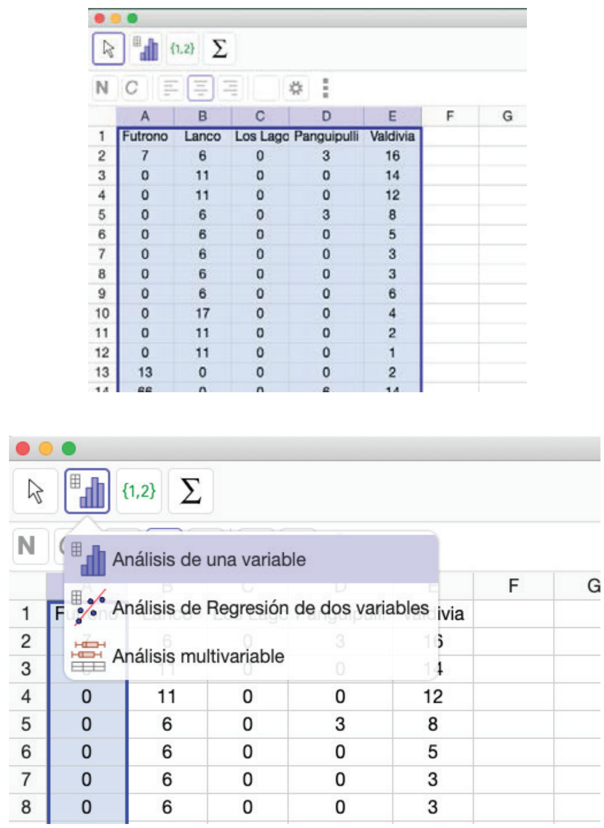


Figura 20 . Vista con hoja de cálculo en GeoGebra Clásico 6.0
Nota. Elaboración propia.

Luego de configurar la vista de GeoGebra, se toma la información desde la planilla de cálculo y se inserta en la hoja de cálculo, como lo indica la Figura 21.



	A	B	C	D	E
1	Futrono	Lanco	Los Lago	Panguipulli	Valdivia
2	7	6	0	3	16
3	0	11	0	0	14
4	0	11	0	0	12
5	0	6	0	3	8
6	0	6	0	0	5
7	0	6	0	0	3
8	0	6	0	0	3
9	0	6	0	0	6
10	0	17	0	0	4
11	0	11	0	0	2
12	0	11	0	0	1
13	13	0	0	0	2

Figura 21. Hoja de cálculo comunas de RLR (arriba) y selección análisis de una variable (abajo)
Nota. Elaboración propia.

Análisis de una variable

Ejemplificamos la comuna de Futrono para realizar un análisis de una variable. Luego de seleccionar la herramienta de análisis de una variable (Figura 21), en la pantalla aparecerá un histograma asociado a la comuna de Futrono, tal como lo indica la Figura 22.

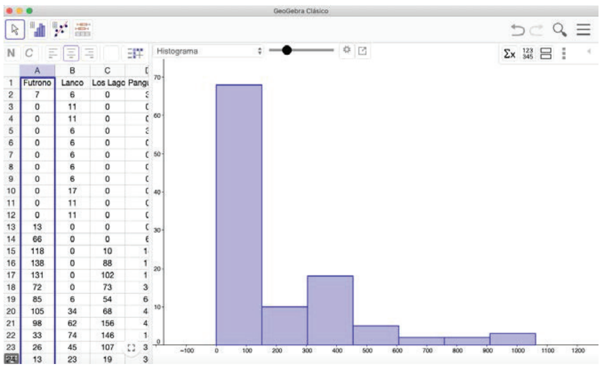


Figura 22 . Histograma por defecto, comuna de Futrono

Nota. Elaboración propia.

Si se desea mostrar otro tipo de representación, se puede seleccionar el siguiente menú. Es posible hacer un diagrama de barras, de caja, de puntos, tallo y hoja y Q-Q plot. Esto siempre dependiendo del tipo de variable que estamos explorando (Figura 23).

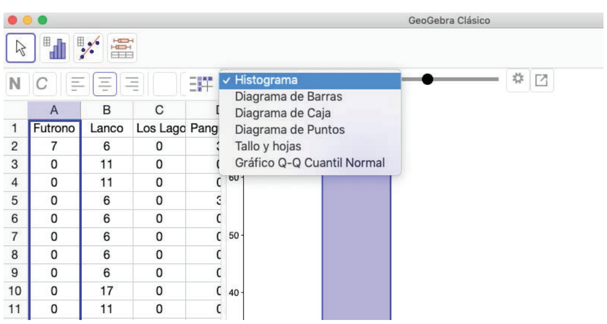


Figura 23. Opciones del tipo de representaciones que brinda GeoGebra

Nota. Elaboración propia.

Al mismo tiempo, si se presiona el icono con forma de engranaje, se puede ajustar el ancho de los intervalos, como lo indica la Figura 24, habilitando algunas opciones para las clases, el tipo de frecuencia y las opciones de habilitar histogramas y polígonos de frecuencias. También se pueden habilitar las estadísticas asociadas a la comuna de Futrono, como lo indica la Figura 25, donde se muestran otras opciones disponibles, en ellas podemos ver que se pueden realizar algunos test de hipótesis y construir intervalos de confianza.

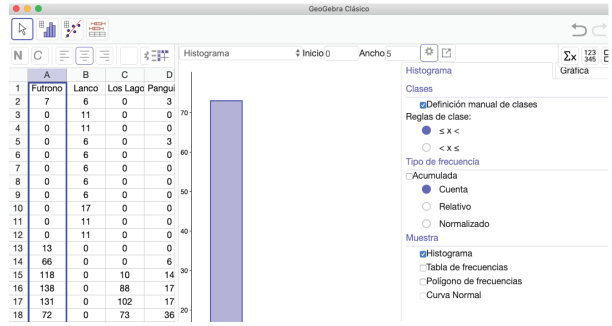


Figura 24 . Icono de configuración de elementos de un histograma

Nota. Elaboración propia.

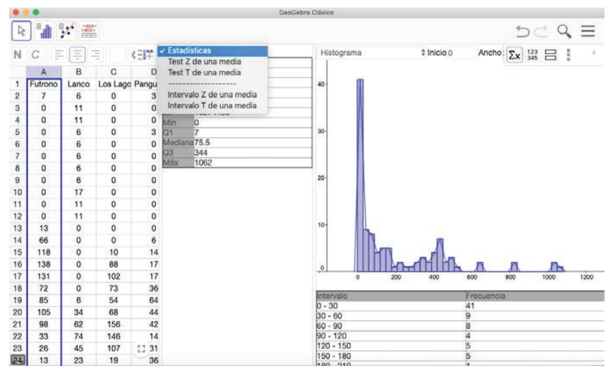


Figura 25. Otras opciones de la herramienta estadísticas

Nota. Elaboración propia.

Análisis de regresión con GeoGebra

Para la construcción de los gráficos de regresión usando GeoGebra Clásico 6.0, debemos seguir los siguientes pasos. Con la información de la planilla de cálculo de las comunas de la RM, seleccionamos dos comunas, para efectos del ejemplo Puente Alto y Vitacura, y presionamos en el menú correspondiente (Figura 26).

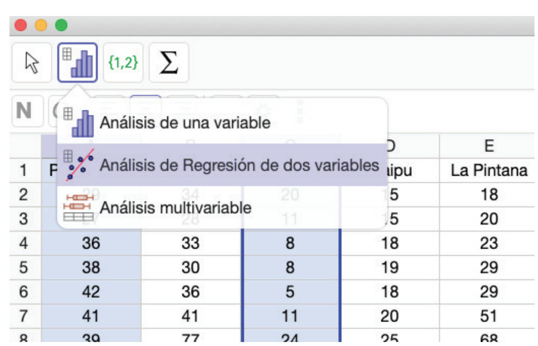


Figura 26. Selección de variables para regresión y agrupamiento

Nota. Elaboración propia.

Cuando se genera el gráfico de dispersión (Figura 27) se habilitan además algunos iconos como el de las estadísticas, visualización de puntos en caso de querer borrar –por ejemplo, algún dato fuera de línea– y un botón que nos permite cambiar de eje la variable, en este caso las comunas de Puente Alto y Vitacura.

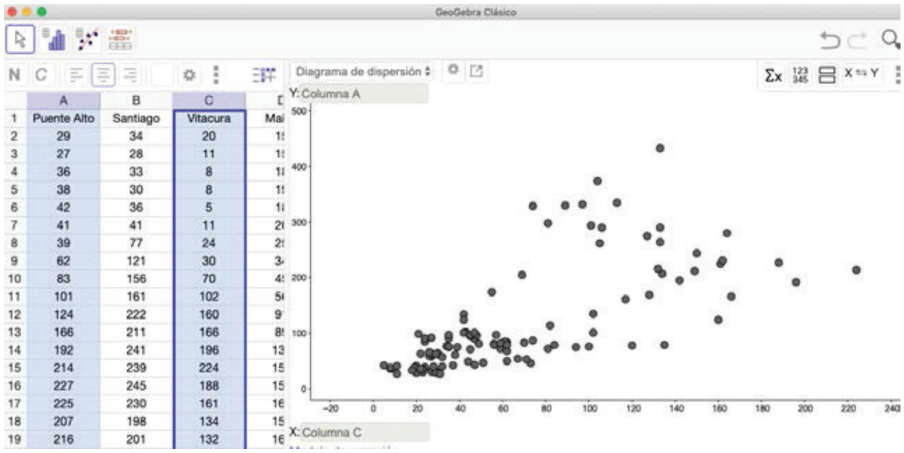


Figura 27. Regresión y agrupamiento para comunas de Puente Alto (columna A - eje Y) y Vitacura (columna C - eje X)

Nota. Elaboración propia.

En la parte inferior de la pantalla aparece un menú que nos permitirá seleccionar algún modelo de ajuste: lineal, log, polinomio (hasta grado 9), entre otros. Para esta propuesta seleccionaremos el modelo de regresión lineal (Figura 28).

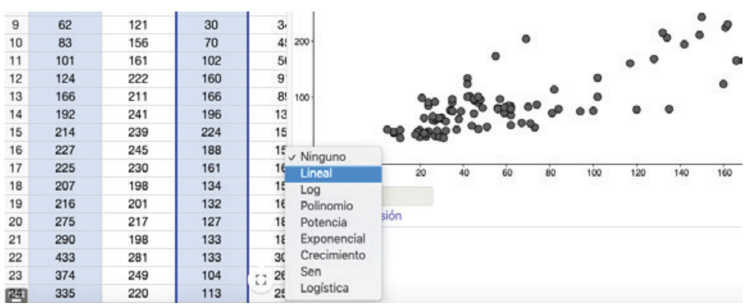


Figura 28 . Selección de modelos de regresión Puente Alto (columna A - eje Y) y Vitacura (columna C - eje X)

Nota. Elaboración propia.

Cuando seleccionamos el modelo de regresión lineal, GeoGebra nos muestra de inmediato en la pantalla la línea de tendencia (color rojo) y junto con esto entrega en la parte inferior la ecuación que representa dicho modelo, el cual por lo tanto nos permite predecir asignando un valor a la variable X. Junto con esto, en la parte superior presionamos el icono de estadísticas, el cual nos genera algunos estadísticos relevantes que pueden ser parte en un análisis exploratorio, estos son medias, desviaciones estándar y varianzas de X e Y. También entrega coeficiente de correlación, covarianza, el valor de ρ (rho) y coeficiente de determinación, R^2.

Análisis multivariado con GeoGebra

Con la información de la planilla de cálculo de la Figura 18, seleccionamos todas las comunas y pinchamos en el menú, seleccionando la herramienta análisis multivariable, como lo indica la Figura 29. A continuación, al costado derecho se generan los gráficos de cajas para cada comuna.

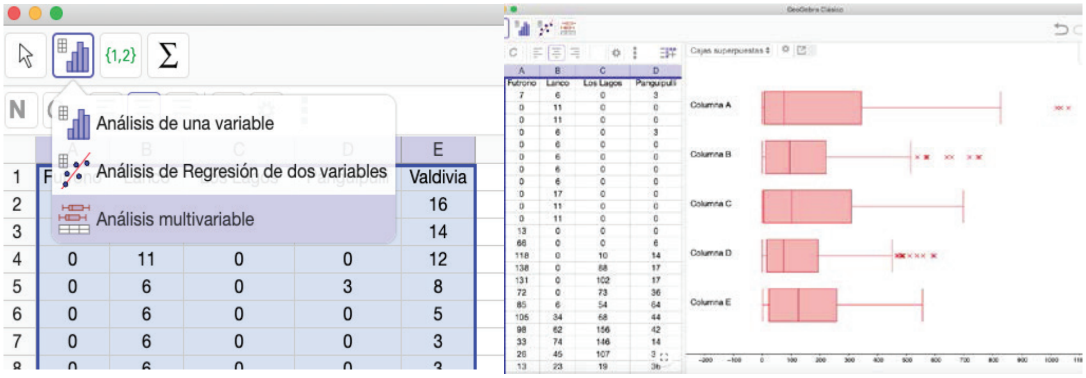


Figura 29 . Selección de comunas para un análisis multivariable (izquierda) y gráficos de caja para las comunas RLR (derecha) Nota. Elaboración propia.

Al hacer clic en el botón ubicado en el costado derecho se pueden habilitar las estadísticas asociadas a cada comuna. Además, en el menú ubicado en la parte inferior de la ventana (Figura 30), se encuentran algunas herramientas para calcular ANOVA, test de diferencias e intervalos de confianza de medias y pareadas.

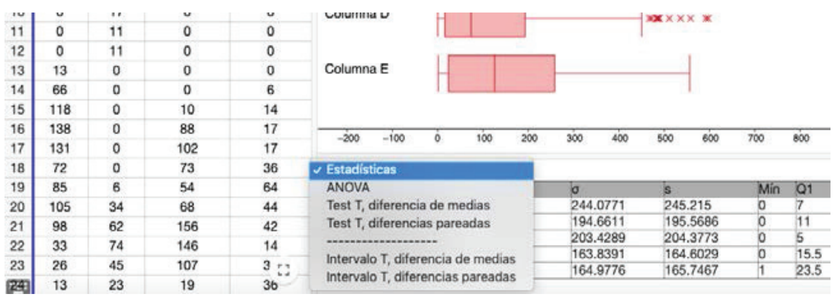


Figura 30. Otras herramientas desde el análisis multivariable
Nota. Elaboración propia.

SOCIEDAD CHILENA DE EDUCACIÓN MATEMÁTICA
REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA
ISSN 2452-5448

ARTÍCULOS DE INVESTIGACIÓN

https://doi.org/10.46219/rechiem.v13i4.86

INTERPRETACIÓN Y COMPRENSIÓN DE GRÁFICOS ESTADÍSTICOS POR PROFESORES DE MATEMÁTICA EN FORMACIÓN

INTERPRETATION AND UNDERSTANDING OF STATISTICAL GRAPHS BY MATHEMATICS STUDENT TEACHERS

Nicolás Sánchez Acevedo
nsanchez@uahurtado.cl
Universidad Alberto Hurtado, Santiago, Chile

Elizabeth Toro Barbieri
elizabethandreatb@gmail.com
Universidad Alberto Hurtado, Santiago, Chile

Daniela Araya Bastias
daniela.araya@ucentral.cl
Universidad Central de Chile, Santiago, Chile

RESUMEN

La importancia que tienen los gráficos estadísticos hoy en día es incuestionable dada la alta presencia en diversos medios para representar información. Con la intención de poder complementar la investigación en el área, llevamos a cabo un trabajo de replicación con profesores de Matemática en formación. A partir de esto, se realizó un trabajo con 23 estudiantes de la carrera de Pedagogía en Matemática, quienes cursan en su mayoría el 6o semestre de formación, para explorar en la comprensión gráfica que estos presentan en una actividad sobre esperanza de vida. El estudio fue realizado bajo una metodología cualitativa, haciendo uso de la técnica de análisis de contenido. Dentro de los resultados más significativos, se logró evidenciar una alta clasificación en los niveles 0 de perspectiva personal que se explica en el presente estudio, y nivel 1, de lectura de datos, mostrándose en algunas interpretaciones errores de lectura. A partir de estos resultados, se hace necesario que las casas de estudio puedan reformular las habilidades en estadística que desarrollen y profundicen aspectos conceptuales de la estadística en la formación de los profesores de Matemática, para promover una enseñanza de la estadística que implique una comprensión profunda de esta disciplina.

PALABRAS CLAVE:
Profesores de Matemática, Cualitativo, Comprensión gráfica, Gráficos estadísticos.

ABSTRACT

The importance of statistical graphs today is unquestionable given the high presence in various media to represent information. With the intention of complementing research in this area, we carried out a replication work with Mathematics student teachers. The study was carried out with 23 students from a Mathematics Education program, who are mostly in the 6th semester of training and explored the graph understanding demonstrated in an activity on life expectancy. The study was carried out using a qualitative methodology, applying the content analysis technique. Among the most significant results, it was possible to show a high classification in levels 0 of personal perspective which is explained in the present study and, level 1 of data reading, showing reading errors in some interpretations. Based on these results, it is necessary that education institutions reformulate the skills in statistics so they will develop and deepen conceptual aspects of statistics in teacher training and promote a teaching of statistics that includes a deep understanding of this discipline.

KEYWORDS:
Math teachers, Qualitative, Graph understanding, Statistical graphs.

Recibido: 5 de junio de 2021, Aceptado: 10 de Agosto de 2021

1. Introducción

Hoy en día, la importancia que tiene el desarrollo de la cultura estadística en la formación de ciudadanos críticos es incuestionable, más aún, considerando la gran cantidad de información que emana de diferentes medios. También esta relevancia se ha acentuado en contextos de enseñanza, tanto en el primario como en el secundario, ganando un trascendental reconocimiento como un componente importante en diferentes planes curriculares de matemáticas (e. g. Ministerio de Educación, Cultura y Deporte de España [MECD], 2015; Ministerio de Educación de Chile [MINEDUC], 2015; National Council of Teachers of Mathematics [NCTM], 2000).

Gran parte de las decisiones oficiales que se incluyen en el currículo y que se aplican en el sistema educativo, muchas veces no responden a lo que la sociedad necesita (Batanero, 2002), por ejemplo, Holmes (2002) plantea que muchas veces las lecciones y actividades que se proponen en documentos curriculares (programas de estudio y libros de texto) son formuladas por matemáticos. Este aspecto, de cierta manera, limita el desarrollo de una adecuada alfabetización estadística, como una habilidad primordial para comprender variada información y poder tomar decisiones con base en los datos disponibles (Meletiou-Mavrotheris y Stylianou, 2003). En este mismo sentido, la toma de decisiones no es una tarea trivial, dado que mucha de la información que es presentada en diversos medios de comunicación es exhibida de múltiples formas, lo que en algunos casos puede generar dificultades al momento de analizar e interpretar dicha información (Fitzallen, 2016; Gea et al., 2017; Meletiou y Lee, 2002).

Tal como plantean Shaughnessy et al. (1996) y Pfannkuch (2006), los gráficos en estadística son usados como representaciones de datos, los que permiten establecer razonamientos para poder comprender y profundizar en el contexto de una situación particular, generar nueva información o aprender de los datos, así como también para lograr indagar en la comprensión de la naturaleza y el tipo de razonamiento que está involucrado, por ejemplo, al momento de dar sentido a los datos, al conjeturar sobre las tendencias, o cuando se realizan posibles inferencias hacia la población.

Dentro de la amplia cantidad de herramientas de las que se disponen para realizar análisis estadísticos, los gráficos estadísticos son parte relevante de una adecuada cultura estadística (Díaz-Levicoy, 2018). En esta misma línea, Schield (2006) plantea que no basta solo con una lectura literal, sino que se hace necesario una adecuada cultura estadística para visualizar tendencia y variabilidad en los datos, pues, por ejemplo, los gráficos circulares y de barras se utilizan especialmente para presentar, difundir y explicar información en los medios (Meletiou-Mavrotheris y Stylianou, 2003).

Para lograr lo anterior, se hace necesario que los profesores, tanto en ejercicio como en formación, tengan desarrolladas las habilidades mínimas y necesarias para poder promover la comprensión de representaciones gráficas en sus estudiantes, siendo una componente necesaria de la cultura estadística del ciudadano actual (Arteaga et al., 2011; Del Pino y Estrella, 2012). Para ello, en el caso de los profesores en formación, estos requieren ser capaces de “motivar la recolección y estudio de datos y de conducir el aprendizaje de las herramientas básicas de su representación y análisis” (Centro de Estudios de Políticas y Prácticas en Educación [CEPPE], 2012, p. 21), que es uno de los estándares disciplinarios en la formación de profesores de Matemática de educación media (CEPPE, 2012).

La investigación sobre comprensión gráfica ha ido en aumento, dada la relevancia que cobra como elemento de la cultura estadística (Gal, 2002). En este mismo sentido, es relevante destacar que, según Gal y Murray (2011), la cultura estadística es parte integrante de dos componentes, una de ellas es la evaluación e interpretación crítica de la información, y la otra es la formulación y comunicación de opiniones respecto a diversa información. Algunas de estas investigaciones han tenido foco tanto en profesores de secundaria como de primaria (Gea et al., 2017; Pfannkuch, 2007), y otras más se han centrado en estudiantes en contexto escolar (Bolch y Jacobbe, 2019; Díaz-Levicoy y Arteaga, 2019).

Por ejemplo, el trabajo de Carmona y Cruz (2016) tuvo por propósito estudiar e identificar aspectos que promueven la comprensión de estudiantes de 7° grado en Colombia en relación con la información que se presenta en gráficos y tablas estadísticas. El análisis se lleva a cabo a partir de la jerarquía de Aoyama (2007). Dentro de los resultados encontrados, se aprecia que los niveles de lectura sobre gráficos están centrados principalmente en los niveles idiosincrático (nivel 1) y lectura básica (nivel 2).

Por su parte, Díaz-Levicoy (2018) llevó a cabo una investigación sobre la comprensión de gráficos estadísticos en alumnos de primaria en Chile. Para ello se apoyó en el enfoque ontosemiótico, realizando un análisis curricular, complementando con la forma en que comprenden gráficos estadísticos estudiantes de 6° y 7° básico. Dentro de los hallazgos más relevantes se encuentra que los estudiantes, al finalizar su educación primaria, presentan dificultades cuando realizan gráficos estadísticos y un conocimiento de lectura bajo en relación con las propuestas curriculares.

García-García et al. (2020) realizaron una investigación en secundaria, que tuvo por objetivo analizar los niveles de comprensión gráfica de estudiantes a partir de tres actividades con gráficos: circular, de barras y líneas. Se apoyaron en los niveles de comprensión gráfica de Curcio (1989) y Friel et al. (2001) y la jerarquía de Aoyama (2007), los cuales articulan para profundizar en la lectura gráfica. Dentro de los resultados encontrados se muestra que una alta cantidad de estudiantes de secundaria alcanza el nivel 2 de comprensión gráfica de manera correcta, y en menor cuantía, algunos estudiantes logran realizar interpretaciones comparativas, en la lectura e interpretación del gráfico circular y de barras. En niveles de mayor lectura se destacan las respuestas de algunos estudiantes que conectan el contexto con interpretaciones.

Vásquez (2021) conduce un estudio con el objetivo de analizar las respuestas de profesores de primaria en Chile para evaluar competencias de sostenibilidad. Para ello utilizó gráficos estadísticos que provenían de diferentes medios de comunicación en el contexto COVID-19. Dentro de los resultados propuestos, se evidencia que es necesario potenciar elementos de conocimiento en relación con la comprensión de gráficos estadísticos y el cómo los profesores hacen uso de estos gráficos para educar en la sostenibilidad.

A partir de lo anterior, y para complementar los aportes en la línea de la comprensión gráfica, hemos planteado esta investigación como un estudio de replicación sobre el trabajo realizado por Gea et al. (2017, en adelante estudio base). Este estudio buscó evaluar la interpretación de gráficos estadísticos por profesores que cursaban el máster en formación del profesorado en secundaria, considerando tres tipos de gráficos (ojiva, gráfico de cajón e histograma) y una tabla con estadísticos de forma complementaria para apoyar las interpretaciones de los estudiantes. En este caso, optamos por un estudio de replicación con la idea de ampliar en la comprensión sobre interpretación y descripción de gráficos estadísticos que llevan a cabo futuros profesores de Matemática, pues si bien es cierto que se ha extendido la investigación en la comprensión gráfica, en el contexto de profesores en formación sigue siendo escasa, y un trabajo de este tipo permitiría aportar en esta línea de trabajo en un contexto poco explorado.

En este sentido, y con base en lo que nos proponemos en este trabajo, adherimos a la idea propuesta por Sánchez (2020), quien plantea que este tipo de estudios “nos permite comprender con mayor profundidad algunos de los fenómenos que hemos identificado y estudiado en el campo de la educación matemática” (p. 39). Además, permite extender la comprensión de las variables de contexto que pueden o no incidir en los hallazgos de investigación (Cai et al., 2018), en nuestro caso, bajo la hipótesis de que estudiantes para profesor de Matemática de cuarto año (que cursaban su tercer año de carrera en 2° semestre de 2020), debiesen movilizar niveles de comprensión gráfica de mayor elaboración y conexión entre conceptos estadísticos. Esto se apoyaría bajo el supuesto que, al momento de llevar a cabo la investigación, los profesores ya habían realizado los cursos de estadística descriptiva, probabilidades e inferencia estadística.

En el caso de nuestra investigación, es un estudio de replicación cerrada, pues nos ajustamos, tanto a los procedimientos del estudio base, es decir, conservamos el marco teórico y metodológico para llevar a cabo la recolección de datos y su posterior interpretación (Hüffmeier et al., 2016). Además, dadas las características que se deben cumplir para realizar este tipo de investigación y asegurar su objetividad, certeza, credibilidad y generalización (Sánchez, 2020), seguimos la forma *scaling out* (Sánchez, 2020). El *scaling out*, y con base en el tipo de exploración que se puede llevar a cabo (Sánchez, 2020), consiste en replicar un estudio, pero haciendo uso de una población con características que difieren de las planteadas en el estudio original. En este caso, la población del estudio original eran estudiantes que participaban de un máster de formación de profesorado en España, y el estudio trabajó con una población de profesores de Matemática en formación en un curso de Didáctica de la Estadística y Probabilidad.

1.2 Objetivo de la investigación

Dados los aspectos descritos en el apartado anterior, y bajo la perspectiva que tiene este trabajo de replicación, nos hemos propuesto como objetivo explorar y describir los niveles de comprensión gráfica que logran futuros profesores de Matemática en una universidad privada en Chile al leer gráficos estadísticos. Para lograr el objetivo planteado, nos apoyamos en los niveles de compresión gráfica propuesto por Friel et al. (2001), que detallamos en el siguiente apartado.

2. Fundamento teórico

La creciente información que es presentada en diferentes medios de comunicación ha llevado a prestar atención en la forma en cómo se construyen y analizan gráficos estadísticos en diversos campos. La capacidad para analizar e interpretar información estadística de manera crítica y poder reflexionar y comunicar ideas es parte de la cultura estadística que todo ciudadano debiera poseer. En particular, ser capaz de leer e interpretar diferentes gráficos estadísticos es componente fundamental de esta alfabetización, pues gran parte de la información que todo ciudadano hoy en día observa en su vida cotidiana se muestra en tablas estadísticas o gráficas (Bolch y Jacobbe, 2019).

La comprensión gráfica no se centra solamente en la lectura e interpretación de gráficos (a secas), sino que incluye otros aspectos que permiten profundizar en ellos, como también abordar la construcción e invención de gráficos en diversos contextos (Friel et al., 2001). Ya sea que la comprensión de gráficos se realice de forma escrita o simbólica, implica tres diferentes comportamientos: traducción, interpretación y extrapolación/interpolación.

La traducción es la capacidad que se puede tener para interpretar diversos datos en gráficos y tablas estadísticas a nivel descriptivo. La interpretación tiene como foco realizar una ordenación y reorganización de la información en las componentes de un gráfico, considerando una relación entre los aspectos más relevantes a los menos relevantes, y la extrapolación/ interpolación implica una extensión de la interpretación, es decir, implica identificar tendencias y posibles consecuencias derivadas de la información gráfica.

En este sentido, la interpretación de un gráfico estadístico conlleva realizar una traducción entre el estudio (contexto) y su representación en el gráfico, es decir, se plantea que no debe estar construido de forma aislada. Díaz-Levicoy (2018) plantea que un gráfico es un sistema semiótico complejo, que requiere primero de la interpretación de cada elemento que lo constituye de forma separada, y luego de manera conjunta.

Aun cuando la comprensión de gráficos y su interpretación no es una tarea trivial, Friel et al. (2001, citando a Bright y Friel, 1996, 1998) han planteado algunos de los beneficios que tiene por ejemplo la transición entre gráficos particulares para promover la comprensión. Por ejemplo, si el interés es resaltar las relaciones estructurales, es posible que, desde un gráfico de barras, que muestra las frecuencias agrupadas, pueda ser más simple transformar a un gráfico de líneas para resaltar similitudes y diferencias entre estas dos representaciones.

Partimos de la base que transiciones particulares entre gráficos permiten beneficios específicos para promover la comprensión gráfica, como la integración conceptual, lectura profunda de gráficos, interpretaciones, inferencias, etc. Además de las transiciones entre gráficos se puede sumar la transición entre tablas y gráficos. Algunas de estas transiciones permiten resaltar las relaciones de estructura entre diversos gráficos, como, por ejemplo, un histograma, una ojiva, un gráfico de cajón y un diagrama de tallo y hoja de forma conjunta (Friel et al., 2001).

A partir de lo anterior, en este trabajo nos apoyamos en el marco de los niveles de lectura (Curcio, 1989; Friel et al., 2001), el cual nos permitirá examinar la forma de lectura y cómo interpretan diversos gráficos estadísticos profesores de Matemática en formación.
En una primera instancia Curcio (1989) plantea tres niveles de comprensión de lectura para gráficos estadísticos, estos son:

Nivel 1. Leer los datos. Este nivel hace referencia a una lectura textual o literal de la información que se representa en una gráfica estadística. Por ejemplo, en un gráfico de cajón identificar el valor de la mediana, o en un histograma, identificar la frecuencia (relativa o absoluta) asociada a un intervalo en el eje de las ordenadas.

Nivel 2. Leer dentro de los datos. Este nivel hace referencia a alguna característica que no es explícita en el gráfico estadístico, es decir, implica un grado de conocimiento y comprensión de conceptos estadísticos. Este nivel también hace necesario la aplicación de procedimientos matemáticos (extrapolación, operaciones simples, comparaciones, etc.). Por ejemplo, en un gráfico de cajón, encontrar el rango intercuartílico (como medida de dispersión), que además de identificar el valor del primer y tercer cuartil requiere de una sustracción para obtener esta medida.

Nivel 3. Leer más allá de los datos. Este nivel se refiere a la obtención de información que no es explícita, que no está representada en el gráfico estadístico y que no es posible analizar con operaciones matemáticas simples ni comparaciones. Por ejemplo, establecer la relación que hay entre la simetría que se presenta a través de un histograma y su relación con características de simetría o asimetría derivadas de comparar la media aritmética, mediana y moda (símbolos). También, a partir de un gráfico cualquiera, inferir o extrapolar alguna conclusión fuera del marco de los datos a la mano.

Años más tarde, Friel et al. (2001) proponen un cuarto nivel de lectura, que hace necesario poseer un conocimiento contextual del cual emerge la gráfica, lo que permite integrar mayores características para la comprensión de gráficos:

Nivel 4. Leer detrás de los datos. Este nivel permite hacer una valoración de forma crítica de la información que se presenta en una gráfica estadística, profundizar en las conclusiones, las que se deben asociar al conocimiento desde donde emerge la problemática, como también la forma de obtener conclusiones y la información (los datos). En este nivel, se es capaz de integrar el conocimiento contextual con las conclusiones, es decir, las conclusiones tienen una base fundamentada en el problema desde donde se extraen los datos y cómo se recolectaron. Un ejemplo de este nivel es hacer un análisis crítico del gráfico en relación a la obtención de los datos, el contexto, el modo de análisis o la crítica del tipo de muestreo y su comportamiento.

Hemos optado por este modelo y sus cuatro niveles, pues nos permite realizar una caracterización y evaluación sistemática de la comprensión sobre gráficas estadísticas por futuros profesores de Matemática.

3. Método

En nuestra investigación, nos apoyamos en un enfoque de tipo cualitativo (Vasilachis, 2006), pues buscamos la comprensión profunda de un fenómeno educativo específico y poder generar las bases para realizar transformaciones de las prácticas en un escenario socioeducativo particular; y de tipo descriptivo (Hernández et al., 2010).

La técnica para analizar las respuestas a la actividad planteada que hacen los profesores de Matemática en formación es el análisis de contenido, debido a que nos permite buscar y explorar en diferentes medios, como documentos escritos, notas de campo, entrevistas etc. (Zapico, 2007). Además, este tipo de técnica se muestra como "objetiva y sistemática [...] que trabaja con materiales representativos, denotados por la exhaustividad en este tipo de técnica" (Porta y Silva, 2003, p. 8).

3.1. Contexto y profesores participantes

En esta investigación participaron 23 estudiantes que cursaban la carrera de Pedagogía en Matemática una universidad privada sin fines de lucro en Chile. Al momento de realizar investigación, el grupo de estudiantes se encontraba cursando la cátedra de Didáctica de la Estadística y las Probabilidades que se dicta en 8° semestre de la carrera. Los estudiantes participantes fueron seleccionados por medio de un muestreo no probabilístico por conveniencia, pues eran estudiantes de cuarto año de la carrera. Los estudiantes pertenecientes al curso mencionado habían aprobado las cátedras de Estadística Descriptiva, dictada en el primer semestre, Probabilidades y Laboratorio de Estadística y Probabilidades, y Estadística Inferencial al momento de estar cursando Didáctica de la Estadística.

Además de ser un trabajo formativo dentro de las actividades de evaluación del curso, a los estudiantes se les informó sobre su participación en este trabajo, el objetivo de la investigación y el uso anónimo que tendrían los resultados derivados de sus respuestas, junto a los fines de mejora continua dentro de la carrera.

3.2. El taller y la tarea propuesta

La actividad que se presentó a los profesores de Matemática en formación (Figura 1) fue tomada de Gea et al. (2017), dado que era una actividad probada en contexto similar al nuestro (formación de profesores), además de encontrarse inserta bajo las mismas temáticas que se desarrollaban en el curso de Didáctica de la Estadística y Probabilidades. El taller que se propuso a los estudiantes forma parte de una serie de actividades curriculares (talleres) que los futuros profesores debían desarrollar con la finalidad de contrastar aspectos teóricos del curso y reflexionar sobre la forma en que ellos y ellas estaban pensando la estadística y las propuestas que emergen de la comunidad de educadores estadísticos. Tal como en Gea et al. (2017), el objetivo de la actividad fue que los estudiantes estudiarán los datos sobre la variable estadística "esperanza de vida al nacer", de 193 países. Los datos fueron obtenidos de la base de las Naciones Unidas (http://hdr.undp.org/es/data).

La actividad estuvo compuesta por cuatro tipos de representaciones para la variable *esperanza de vida*. Esta incluyó un histograma de la distribución de la variable, una ojiva de frecuencias acumuladas, un gráfico de cajón y una tabla resumen con los estadísticos básicos de la variable.

Esperanza de vida. Hemos preparado una tabla de frecuencias, gráficos, además de una tabla con las estadísticas básicas sobre la esperanza de vida de un total de 193 países. Interpreta cada uno de estos gráficos, explicando qué nos indican sobre la distribución de la esperanza de vida.

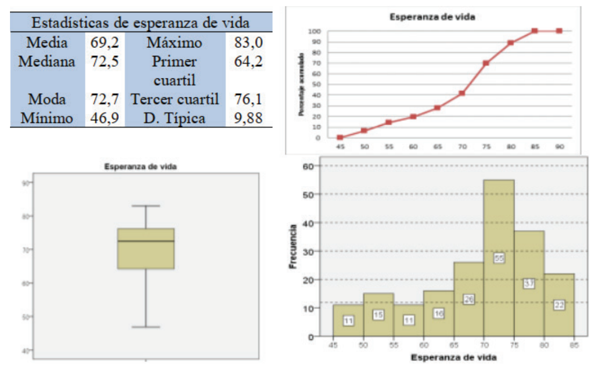


Figura 1. Gráficas estadísticas y tabla de resumen para la variable esperanza de vida. Nota. Elaboración propia a partir de la base de datos de las Naciones Unidas (http://hdr.undp.org/es/data).

Consigna

Interprete cada una de las tres gráficas estadísticas que se presentan para explicar las características que se pueden deducir de la distribución de la *variable esperanza de vida*. Puede apoyar la interpretación a partir de la tabla "Estadísticas de la esperanza de vida".

3.3. Análisis de los datos

Para realizar la clasificación y análisis de las respuestas que entregaron los futuros profesores de Matemática en relación con sus niveles de comprensión gráfica (Curcio, 1987; Friel et al., 2001), consideramos categorías, que son los niveles de comprensión gráfica (leer los datos, leer dentro de los datos, leer más allá de los datos y leer detrás de los datos), y subcategorías, que están relacionadas con el tipo de interpretación que hacen del gráfico propuesto y posibles conexiones. Además de los cuatro niveles de comprensión gráfica, hemos considerado un *Nivel 0*: perspectiva personal, nivel que muestra que la lectura es desde una perspectiva personal o la propia experiencia sin establecer conexiones (Fernández et al., 2019; García-García et al., 2020). Un ejemplo de interpretación es el que mostramos en la Figura 2:

> -el grafico de bigotes aporta de manera visual al comportamiento de datos que se expresan en la tabla, se puede observar de mejor manera la distribución de los datos que aparecen en la tabla, así como saber su asimetría.

Figura 2. Ejemplo de respuesta de estudiante. Nota. Respuesta dada por estudiante 15 (E-15) a la consigna propuesta sobre el gráfico de cajón.

Vemos que el nivel de lectura es nulo (nivel 0), dado que no hace interpretación explícita de la gráfica, es decir, no se aprecia ni siquiera una lectura literal del gráfico de cajón. En relación con la subcategoría, esta es la descripción que se hace de la gráfica, que en este caso viene dada al hacer una interpretación enfatizando la utilidad, es decir, su interpretación está en describir para qué sirve el gráfico de manera general.

Para la homogeneización de las respuestas y su clasificación, seguimos un proceso de tipo inductivo. En cuanto a la fiabilidad y rigurosidad del análisis efectuado, cada uno de los investigadores realizó un análisis independiente que comprendió los siguientes pasos: (i) analizar las respuestas de los profesores en formación y clasificar con base en los niveles de comprensión gráfica, (ii) justificar y argumentar la decisión de situar las respuestas en determinado nivel (esto se hizo de manera escrita), (iii) describir y clasificar las respuestas de los profesores de acuerdo a alguna subcategoría, incluyendo si la respuesta era correcta o incorrecta en la interpretación, y finalmente (iv) se realizó una discusión conjunta entre los investigadores sobre la clasificación e interpretación de categorías y subcategorías, y en la situación de tener divergencias en los análisis independientes, se volvió a discutir la respuesta dada por los futuros profesores para consensuar la clasificación e interpretación.

4. Análisis de los resultados

En este apartado, mostramos los resultados dados por las evidencias en las respuestas de los profesores en formación en relación con la tarea que se presentó en el apartado 3.2. Las respuestas se analizaron textualmente de las hojas entregadas por los profesores (en formato imagen, dado que en contexto de pandemia enviaron sus trabajos en .pdf y/o .jpeg), con la intención de no omitir detalles en la descripción de la lectura e interpretación y que estas fueran un fiel reflejo de la realidad. En algunos casos, fue necesario realizar transcripción de algunas respuestas, dado que presentaban errores de redacción o la tipología no era clara.

Los resultados se presentan en dos apartados, para dar mayor claridad en la comprensión gráfica de los profesores. En el primer apartado (4.1.), se muestran las respuestas entregadas por los profesores en relación con la interpretación de cada uno de los gráficos estadísticos, las cuales se clasificaron de acuerdo a las categorías propuestas y que definimos como los niveles de comprensión (leer los datos, leer dentro de los datos, leer más allá de los datos, leer detrás de los datos, y la inclusión del "nivel 0").

En el caso de las subcategorías, las consideramos como las estrategias de interpretación de los profesores, tanto correctas como incorrectas.

En el segundo apartado (4.2.), mostramos la distribución de las categorías de las respuestas de los futuros profesores en relación con el tipo de respuesta entregada, si es correcta o incorrecta y su interpretación.

4.1. En relación con los niveles de lectura evidenciados y sus interpretaciones

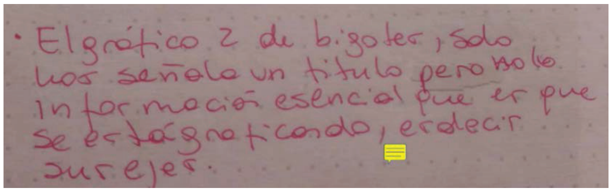
Mostramos en este apartado las interpretaciones que realizan los profesores en relación con la consigna propuesta para la interpretación y lectura de los gráficos estadísticos. Las respuestas están presentadas según niveles de lectura sobre los tres tipos de gráficos (histograma, ojiva y gráfico de cajón), como también interpretaciones correctas e incorrectas.

En esta consigna, los niveles identificados que movilizan los futuros profesores están en un nivel 0 y un nivel 1. Solo uno de los profesores en formación (E-8) logra dar una respuesta que se clasifica en el nivel 2 (lectura dentro de los datos). Algunas de las respuestas que entregan los profesores en formación se muestran a continuación.

4.1.1. Niveles de lectura en relación con el gráfico de cajón

En esta parte del análisis no se encontraron o evidenciaron respuestas que dieran cuenta de una lectura de gráficos que estuvieran en el nivel 2 de interpretación gráfica.

Las respuestas que hacen los profesores en formación en relación con la consigna dos se clasifican en un nivel de compresión gráfica elemental, es decir, "lectura de datos" (nivel 1). Se aprecia que la mayoría de los futuros profesores se centra en describir el gráfico de cajón de manera literal, detallando cuartiles, mediana y/o mínimos y máximos. En algunos casos, la interpretación incluye aspectos de variabilidad, pero solo haciendo mención visual a la diferencia entre cuartiles y mediana. Por otro lado, y en relación con la forma de la distribución, algunos también mencionan la asimetría pero de manera superficial, es decir, no establecen relaciones, por ejemplo, con los estadísticos presentados. Algunas de las interpretaciones las mostramos a continuación:

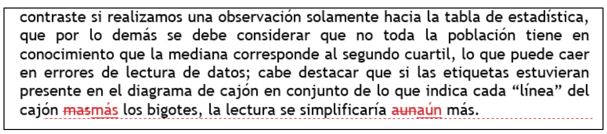


· El gráfico 2 de bigotes, solo nos señala un título pero no la información esencial que es que se está graficando, es decir sus ejes.

Figura 3. Respuesta desde una perspectiva personal (nivel 0)

Nota. Respuesta del estudiante 4 sobre el gráfico de cajón.

Vemos aquí que el estudiante no establece una conexión entre la tabla donde se presentan los estadísticos descriptivos y el diagrama de bigotes. El foco en la respuesta de E-4 es hacer mención solo al título y que de este no es posible extraer información relevante, pues sus ejes no están estratificados. Así mismo, no se refiere, en su interpretación, a la variable estadística que es considerada en la actividad (esperanza de vida).

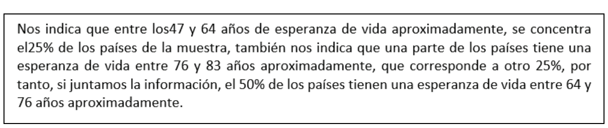


contraste si realizamos una observación solamente hacia la tabla de estadística, que por lo demás se debe considerar que no toda la población tiene en conocimiento que la mediana corresponde al segundo cuartil, lo que puede caer en errores de lectura de datos; cabe destacar que si las etiquetas estuvieran presente en el diagrama de cajón en conjunto de lo que indica cada "línea" del cajón ~~mas~~más los bigotes, la lectura se simplificaría ~~aun~~aún más.

Figura 4. Respuesta bajo una perspectiva personal (nivel 0)

Nota. Respuesta del estudiante 18 sobre el gráfico de cajón.

Vemos de la respuesta entregada por el estudiante 18, que no hace una interpretación del gráfico en relación con la variable esperanza de vida. La respuesta, da cuenta de una visión más instrumentalista del gráfico en términos de la valoración de uso y pertinencia de la gráfica. Además, se interpreta un conocimiento no adquirido sobre la población al interpretar dicho gráfico, de donde emergen errores de lectura. Junto a ello, menciona que la ausencia de etiquetas dificulta la lectura del gráfico, aun cuando los estadísticos y gráficos se presentan de manera conjunta.

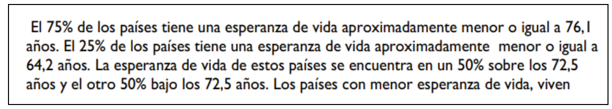


Nos indica que entre los47 y 64 años de esperanza de vida aproximadamente, se concentra el25% de los países de la muestra, también nos indica que una parte de los países tiene una esperanza de vida entre 76 y 83 años aproximadamente, que corresponde a otro 25%, por tanto, si juntamos la información, el 50% de los países tienen una esperanza de vida entre 64 y 76 años aproximadamente.

Figura 5. Respuesta de lectura de datos (nivel 1)

Nota. Respuesta del estudiante 6 a partir del gráfico de cajón.

En este nivel, apreciamos que E-6 hace una lectura literal de los datos en relación con el gráfico de cajón. Interpreta los cuartiles del gráfico a partir de la estimación que hace de estos valores, apoyándose en los valores de la tabla de los estadísticos descriptivos y el gráfico de cajón, aludiendo a los cuartiles y los intervalos (rango intercuartílico) donde se concentran los países con esperanzas de vida por sobre (o bajo) cierto cuartil.

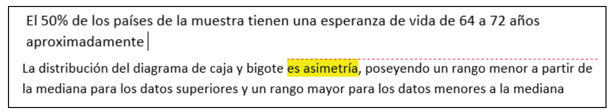


El 75% de los países tiene una esperanza de vida aproximadamente menor o igual a 76,1 años. El 25% de los países tiene una esperanza de vida aproximadamente menor o igual a 64,2 años. La esperanza de vida de estos países se encuentra en un 50% sobre los 72,5 años y el otro 50% bajo los 72,5 años. Los países con menor esperanza de vida, viven

Figura 6. Respuesta en el nivel de lectura de datos (nivel 1)

Nota. Respuesta del estudiante 16 sobre el gráfico de cajón.

En esta respuesta, E-16 interpreta adecuadamente el percentil 75 (cuartil 3) en relación con la variable esperanza de vida, como también lo hace con el percentil 25. La lectura que hace el estudiante es literal, su respuesta se focaliza en describir aquellos elementos distintivos de un gráfico de cajón. No se evidencia relación de la respuesta con la simetría de la distribución.



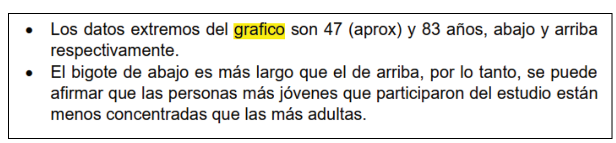
El 50% de los países de la muestra tienen una esperanza de vida de 64 a 72 años aproximadamente

La distribución del diagrama de caja y bigote es asimetría, poseyendo un rango menor a partir de la mediana para los datos superiores y un rango mayor para los datos menores a la mediana

Figura 7. Respuesta en el nivel de lectura de datos (nivel 1)

Nota. Respuesta del estudiante 15 sobre el gráfico de cajón.

La respuesta dada por E-15 nos muestra un nivel de comprensión gráfica en el nivel de lectura literal, en donde vemos que el estudiante interpreta la mediana en función de un intervalo de clase. Este nivel de lectura literal hubiese sido más claro si, por ejemplo, el estudiante hubiese planteado el valor aproximado de la mediana a partir del gráfico de cajón, afirmación que también se hubiese apoyado con el estadístico de la mediana dado en la tabla de estadísticos. Destacamos en su respuesta la mención (informal) del concepto de asimetría, en relación con la forma de la distribución, la cual justifica a partir de la diferencia (que nombra como rango) entre la mediana y los *datos superiores e inferiores*, no especificando si se refiere al máximo y mínimo, los cuartiles 1 y 3, o ambos.



- Los datos extremos del grafico son 47 (aprox) y 83 años, abajo y arriba respectivamente.
- El bigote de abajo es más largo que el de arriba, por lo tanto, se puede afirmar que las personas más jóvenes que participaron del estudio están menos concentradas que las más adultas.

Figura 8. Respuesta en el nivel lectura de datos (nivel 1)

Nota. Respuesta errónea del estudiante 20 sobre el gráfico de cajón.

Si bien es cierto que esta respuesta puede estar dentro de un nivel de lectura literal, presenta un error de interpretación en relación con la variable esperanza de vida. El estudiante hace mención de la edad de las personas y no de la esperanza de vida como variable de estudio. Este error de interpretación se presenta en esta consigna en muy pocos casos.

Tenemos que:

$Q_1 = 64,2$

$Q_2 = 72,5$

$Q_3 = 76,1$

Por lo que observamos que las diferencias entre los cuartiles y la mediana no son mayores, de hecho, es menor que la desviación estándar. Además, la moda se encuentra dentro del rango. Entonces, la distribución de los datos está concentrada entre estos valores. Esto indica que las personas tienden a esperar una vida de entre 64,2 y 76,1 años.

Figura 9. Respuesta de lectura dentro de los datos (nivel 2)

Nota. Respuesta del estudiante 9 con error de interpretación.

Este estudiante da una respuesta que visualizamos que podría estar en el nivel de lectura dentro de los datos, pues se aprecia que explicita cada uno de los cuartiles y la mediana, y a partir de estos (aun cuando no realiza operaciones simples), menciona que hay diferencias entre los cuartiles (Q_1 y Q_3) y la mediana, haciendo alusión a la dispersión. Sin embargo, la respuesta presenta un error en la interpretación de la variable del estudio, pues se refiere a "las personas tienden a esperar una vida" y no a la esperanza de vida. En este nivel, solo logramos incluir (medianamente) la respuesta de este futuro profesor.

4.1.2. Niveles de lectura en relación con la interpretación del histograma y su relación con el gráfico de cajón

En relación con las respuestas que entregan los profesores en esta consigna, la mayoría muestra un nivel de comprensión literal, es decir, de lectura de los datos. Algunos otros profesores en formación focalizan la interpretación del histograma en un nivel de lectura dentro de los datos y muy pocos evidencian una lectura más allá de los datos. A partir de esto, mostramos algunos ejemplos representativos de las respuestas de los profesores.

En relación al histograma creo que es el más correcto para presentar esta información a las personas, en él podemos observar directamente la esperanza de vida la nacer y su frecuencia absoluta, así la gente se puede entender de mejor manera como se distribuye la esperanza de vida al nacer.

Figura 10. Respuesta de un nivel de perspectiva personal (nivel 0)

Nota. Respuesta del estudiante 5 sobre el histograma y gráfico de cajón.

Vemos de esta respuesta que E-5 no realiza ninguna interpretación del histograma, por ejemplo, una estimación de la media aritmética, la forma de la distribución o la simetría. La respuesta hace mención a que el histograma muestra la variable esperanza de vida, incluyendo la frecuencia absoluta, describiendo una relación entre la variable del eje x con la frecuencia en el eje y, sin mayor descripción. Además, establece un juicio de valor mencionando lo adecuado del histograma para presentar la información a las personas.

6 Al igual que el caso anterior se dificulta el interpretar la información si tomamos el eje x como los años vemos que la curva crece por lo que mientras mas años mas esperanzas de vida. que puede llegar a los 90 años.

Figura 11. Respuesta de un nivel de perspectiva personal (nivel 0)

Nota. Respuesta del estudiante 4 sobre el histograma y gráfico de cajón.

La respuesta de esta estudiante es, más bien, una idea general de un histograma. No realiza, en ningún caso, una lectura literal, centrando su explicación en la dificultad de interpretar la información, la que pudo haber complementado con los estadísticos. Alude al crecimiento del histograma, y menciona que la esperanza de vida puede llegar a los 90 años, cuando el máximo es de 83 años.

El histograma muestra la frecuencia, es decir la cantidad de países que tienen su esperanza de vida en distintos intervalos de edad, acá se ve claramente lo mencionado anteriormente en la interpretación del grafico de caja y bigote, la mayor cantidad de países se encuentra en el intervalo entre 70 y 75 años, la mayor cantidad de datos se encuentran hacia la izquierda.

Figura 12. Respuesta en nivel de lectura de datos (nivel 1)

Nota. Respuesta del estudiante 11 sobre el histograma y gráfico de cajón.

En esta respuesta, dada por E-11, vemos que el nivel de comprensión es literal, es decir, es capaz de asociar los ejes con la respectiva representación: el eje y con la cantidad de países, y el x con la esperanza de vida. Además, establece del histograma un intervalo modal para la esperanza de vida de la cantidad de países y, a partir de esto, una escueta idea de la simetría, con sesgo a la izquierda.

La esperanza de vida en 11 países se encuentra entre los 45 y 50 años. Estos países en comparación con el resto son los de menor frecuenta y menor esperanza de vida, esto puede deberse a carencias en el sistema de salud de los países. La esperanza de vida en 55 países se encuentra entre los 70 y 75 años, estos países tienen mayor frecuencia en comparación al resto. En conclusión aproximadamente 140 países tienen una esperanza de vida mayor a 65 años, esto puede deberse a una mejor calidad de vida, y mejor sistema de salud.

Figura 13. Respuesta en nivel de lectura de datos (nivel 1). Nota. Respuesta del estudiante 16 sobre el histograma y gráfico de cajón.

En este caso, apreciamos que la interpretación de E-16 denota un nivel de comprensión literal del histograma, es decir, es capaz de relacionar la variable esperanza de vida (intervalos) con la frecuencia de países en dichos intervalos. Además, de esto, se visualiza una conclusión de causa-efecto que es distintiva, dado que menciona que la causa por la que una gran cantidad de países tienen una esperanza de vida mayor a los 65 años, es debido a una mejor calidad de vida o sistema de salud adecuado.

5) El histograma nos da a conocer facetas muy distintas a las demás, en primera instancia podemos notar que tiene un sesgo hacia la izquierda, por ende, asimetría negativa y eso nos quiere decir tambien que la media es menor que la moda(se ve explícito) de este modo podemos hacer distintas interpretaciones de todo, sobre la moda, la mayor y mejor frecuencia, etc, junto a lo que pudimos ver podemos podemos decir que en base a lo anterior el de cajón tiene

Figura 14. Respuesta en nivel de lectura de datos (nivel 2). Nota. Respuesta del estudiante 19 sobre el histograma y gráfico de cajón.

En esta respuesta podemos ver un nivel de comprensión gráfica que denota una lectura más allá de los datos. E-19 se refiere al sesgo de la distribución hacia la izquierda y lo relaciona con una asimetría de tipo negativa, que implica no solo una lectura literal, sino que también de un conocimiento conceptual que relaciona el orden de los estadísticos (media, mediana y moda). Esto lo señala de manera explícita, cuando dice que la media es menor que la moda.

4.1.3. Niveles de lectura en relación con la ojiva y posibles comparaciones con los otros gráficos

La descripción de la ojiva presenta una gran cantidad de estudiantes que interpretan desde una perspectiva personal, quienes declaran, por ejemplo, explícitamente aspectos sobre el significado de los ejes, como también lo que implica este tipo de gráficos en términos de su uso y la función de acumular frecuencias que, en este caso, representa a la cantidad de países que presentan una esperanza de vida menor (o mayor) a cierta cantidad de años. Otro grupo de estudiantes muestra respuestas que se pueden clasificar dentro del nivel de lectura de datos, dando interpretaciones correctas y una adecuada descripción, siendo literales y refiriéndose a la esperanza de vida en términos acumulados. También se aprecian ciertas descripciones erróneas en la interpretación. No se evidencian respuestas que denotan un nivel de comprensión gráfica que vaya más allá de los datos. Mostramos algunos ejemplos representativos de estos niveles.

La ojiva representa la frecuencia acumulada de los datos, en este caso los datos corresponden a la esperanza de vida de 193 países.
La variable horizontal de la grafica indica las edades de la esperanza de vida que tienen los países, mientras que la variable vertical corresponde a la frecuencia que tiene cada uno de estos datos.

Figura 15. Respuesta en nivel de perspectiva personal (nivel 0) Nota. Respuesta del estudiante 2 en relación con la ojiva.

Dentro de este nivel, vemos que la respuesta de este estudiante (E-2) es específicamente sobre la función que cumple la ojiva, es decir, permite representar frecuencias acumuladas. Es relevante destacar en esta respuesta que el estudiante, si bien en su nivel de comprensión no da cuenta de un nivel literal, considera que ambos ejes representan una variable. Para el eje x, escribe que la variable son las edades de la esperanza de vida, cuando la esperanza está medida en unidad de tiempo, pero para el eje y escribe que se presenta como la variable que corresponde a la frecuencia, denotando un error en relación con el significado de los ejes.

- Un 0% de los países de la muestra, se encuentra su esperanza de vida igual o menor de 45 años.
- Un 5.6% de los países de la muestra, se encuentra su esperanza de vida igual o menor de 50 años.
- Un 13.4% de los países de la muestra, se encuentra su esperanza de vida igual o menor a 55 años.
- Un 19.1 de los países de la muestra, se encuentra su esperanza de vida igual o menor a 60 años.

Figura 16. Respuesta en nivel de lectura de datos (nivel 1) Nota. Respuesta del estudiante 16 en relación con la ojiva.

Dentro de las interpretaciones correctas, la gran mayoría de los estudiantes realiza este tipo de descripción, es decir, asocia numéricamente el eje de la variable esperanza de vida con las frecuencias acumuladas. En este caso, vemos que E-16 realiza una lectura literal de la gráfica de ojiva, dando cuenta de la relación entre la variable esperanza de vida y la frecuencia acumulada. Sí es relevante destacar que el lenguaje utilizado por el estudiante no da cuenta del uso de la variabilidad, pues se desprende de su respuesta un lenguaje determinista al afirmar que un tal porcentaje de países tiene una esperanza de vida específica.

Además, se aprecia un error en la primera interpretación (en consonancia con lo mostrado en Gea et al., 2017) en relación con la interpretación del mínimo valor de la ojiva, declarando que no hay países (0%) con una esperanza de vida menor a 45 años.

encuentran distribuidos, en cambio el grafico de ojiva nos permite ver el porcentaje de países cumplen con la esperanza de vida que se ~~este~~esté observando, entonces podemos entender la gran diferencia de estos como en lo que se enfoca su representación, siendo el grafico de caja un

Figura 17. Respuesta en nivel de lectura de datos (nivel 1) Nota. Respuesta errónea del estudiante 18 en relación con la ojiva.

Respecto a las respuestas erróneas sobre la ojiva, vemos por ejemplo la de E-18, quien realiza una interpretación incorrecta al confundir la frecuencia absoluta acumulada con la frecuencia absoluta. Esto se ve reflejado cuando menciona la esperanza de vida en términos puntuales. El resto de su interpretación se detiene en aspectos generales de la gráfica.

Las interpretaciones del gráfico de ojivas (si estaría correctamente hecho) son muy similares a las de caja y bigotes, ya que solo podemos interpretar bajos los porcentajes que tienen los intervalos, entonces podemos interpretar que el 5% de los países tienen un esperanza de vida de 45-50 años, también que el que los países que tienen una esperanza de vida de 40-75 años son aproximadamente el 70%.

Figura 18. Respuesta errónea en nivel de lectura de datos (nivel 1)
Nota. Respuesta errónea del estudiante 19 en relación con la ojiva.

En esta respuesta, vemos que E-19 describe el gráfico de ojiva comparándolo con el gráfico de cajón, cuando el primero permite estimar frecuencias acumuladas, mientras que en el segundo se dan las otras estimaciones (aproximadas) puntuales. Si bien es cierto que el estudiante logra extrapolar y poner en relación características de los ejes, lo hace erróneamente, pues confunde las frecuencias acumuladas con frecuencias absolutas.

4.2. En relación con la distribución de los niveles de lectura gráfica

En este apartado exponemos los resultados generales por niveles de lectura gráfica evidenciados en las respuestas de los profesores en formación y su ajuste a una adecuada o inadecuada interpretación. Estos se presentan con relación a las respuestas sobre el gráfico de cajón, el histograma y la ojiva.

4.2.1 Interpretación y distribución de respuestas en relación con el gráfico de cajón

Las interpretaciones que hacen los futuros profesores de Matemática en relación con el gráfico de cajón están concentradas mayoritariamente en el nivel 1, muy escasamente algunas en el nivel 0 (perspectiva personal) y solo un estudiante (de acuerdo a nuestra interpretación como investigadores) logra una interpretación en un nivel de lectura que va más allá de los datos. Los resultados se muestran en la Tabla 1.

Tabla 1. Distribución de las respuestas sobre la interpretación del Gráfico de cajón. Nota. NR: No responde.

Distribución de respuestas según nivel de lectura gráfica del gráfico de cajón			
	Correcta	**Incorrecta**	**Totales**
Nivel 0	1 (4,3%)	2 (8,7%)	3 (13,0%)
Nivel 1	17 (73,9%)	1 (4,3%)	18 (78,3%)
Nivel 2	1 (4,3%)	0 (0,0%)	1 (4,3%)
NR	1 (4,3%)		1 (4,3%)
Totales	20 (87,0%)	3 (13,0%)	23

De la Tabla 1, vemos que la mayor cantidad de respuestas en relación con la interpretación del gráfico de cajón está en el nivel de lectura de datos (78,3%), es decir, los profesores en formación logran identificar y relacionar la variable esperanza de vida y algunos de los estadísticos que componen el gráfico de cajón (cuartiles, mediana, máximo y mínimo). Aun cuando el gráfico podía ser interpretado con apoyo de la tabla estadística, las respuestas y un nivel de compresión de mayor profundidad no se evidencian. De las respuestas que se hacen sobre la lectura de datos (nivel 1), un estudiante (4,3%) realiza una interpretación incorrecta, construyendo toda su interpretación a partir de la media aritmética y no de la mediana.

Dentro de las respuestas que hacen los futuros profesores de Matemática al interpretar el gráfico de cajón, mencionan la asimetría de la distribución de la variable esperanza de vida, pero sin entregar un argumento, por ejemplo, se esperaría de ellos que lograsen relacionar la dispersión entre los cuartiles 1 y 3 y la mediana o las distancias entre los cuartiles y los valores máximos y mínimos.

Las respuestas que presentan menor frecuencia (13,0%) son aquellas que no establecen una relación entre los elementos del gráfico de cajón y la variable esperanza de vida. Estas respuestas se centran en describir aspectos de la utilidad del gráfico de cajón. Solo un estudiante (4,3%) no realiza una interpretación de este gráfico presente en la actividad.

4.2.2. Interpretación y distribución de respuestas en relación con el histograma

En relación con la interpretación del histograma se aprecia, tal como en el caso anterior, que la mayor cantidad de estudiantes manifiesta un nivel de lectura literal (nivel 1, 78,3%), y con una menor frecuencia en un nivel de perspectiva personal (nivel 0, 13,0%) y un nivel de lectura que va más allá de los datos (nivel 2, 8,7%). La distribución se muestra en la Tabla 2.

Tabla 2. Distribución de las respuestas sobre la interpretación del histograma de cajón. Nota. NR: No responde.

Distribución de respuestas según nivel de lectura gráfica del histograma			
	Correcta	**Incorrecta**	**Totales**
Nivel 0	1 (4,3%)	2 (8,7%)	3 (13,0%)
Nivel 1	15 (65,2%)	3 (13,0%)	18 (78,3%)
Nivel 2	2 (8,7%)	0 (0,0%)	2 (8,7%)
NR	0 (0,0%)		0 (0,0%)
Totales	18 (78,3%)	5 (21,7%)	23

De la distribución de las respuestas que hacen los profesores en formación sobre la interpretación del histograma, dentro del nivel 1 (lectura de datos) la mayoría logra establecer una relación literal entre la variable esperanza de vida (eje x) y la frecuencia de países en un intervalo determinado. Dentro de este mismo nivel, y en relación con las respuestas correctas, algunas de ellas hacen énfasis (de manera intuitiva) en la asimetría de la distribución, considerando la forma que tiene el histograma, pero sin establecer una relación, por ejemplo, con el orden que puede tener la media aritmética, la mediana y la moda. En ninguno de los casos se habla, por ejemplo, de la dispersión de la distribución, y si en algunas de las respuestas de los estudiantes se hace, es de manera muy somera, sin un argumento que permita validar la interpretación.

Las respuestas que emergen en el nivel de una perspectiva personal (nivel 0) se refieren al uso y utilidad que tiene el histograma para el análisis de información y la interpretación de fenómenos de la vida cotidiana. Dentro de esta misma perspectiva, las interpretaciones que son incorrectas tienen relación con confundir la variable esperanza de vida con números de personas, como también hacer mención del porcentaje de personas y no de la esperanza de vida de los países involucrados en el estudio.

Finalmente, en el nivel de lectura más allá de los datos, dos estudiantes (8,7%) entregan respuestas en este nivel (nivel 2). Las respuestas de estos estudiantes se consideran en este nivel dado que logran establecer en su interpretación una relación entre la asimetría de la distribución, argumentando que la media aritmética es menor que la mediana y la moda, es decir, son capaces de extrapolar información que no es explícita del gráfico.

4.2.3. Interpretación y distribución de respuestas en relación con la ojiva

Las respuestas en relación con la interpretación de la ojiva muestran que la mayor cantidad de estudiantes se encuentra en un nivel de perspectiva personal (52,2%) y en menor frecuencia en un nivel de lectura literal (47,8%), pero con un alto porcentaje (30,4%) de interpretaciones incorrectas. No se evidenciaron respuestas en un nivel de comprensión más allá de los datos (nivel 2). La distribución se presenta en la Tabla 3.

Tabla 3. Distribución de las respuestas sobre la interpretación de la ojiva

Nota. NR: No responde.

Distribución de respuestas según nivel de lectura gráfica de la ojiva			
	Correcta	**Incorrecta**	**Totales**
Nivel 0	10 (43,5%)	2 (8,7%)	12 (52,2%)
Nivel 1	4 (17,4%)	7 (30,4%)	11 (47,8%)
Nivel 2	0 (0,0%)	0 (0,0%)	0 (0,0%)
NR	0 (0,0%)		0 (0,0%)
Totales	14 (60,7%)	9 (39,1%)	23

Dentro de las respuestas que entregan los profesores en formación sobre la ojiva, se aprecia una gran cantidad de dificultades para su interpretación y errores asociados. El 52% de los estudiantes no entrega una interpretación del gráfico en relación con la variable esperanza de vida, sino que se centra en describir aspectos de su utilidad, por ejemplo, mencionando su uso sobre las frecuencias acumuladas. Dentro de las respuestas incorrectas en este nivel, se identificaron interpretaciones inadecuadas sobre el valor máximo acumulado en la ojiva y sobre el valor mínimo, como también interpretaciones intervalares para estimaciones (a partir del gráfico) que pueden haber sido puntuales, más aún, contando con el apoyo de la tabla con los descriptivos.

Para el caso de las interpretaciones en el nivel de lectura de datos, un grupo de estudiantes logra hacer una lectura adecuada, relacionado de forma adecuada la variable esperanza de vida y la frecuencia acumulada en el eje y. Los errores más frecuentes en este nivel se relacionan con confundir la frecuencia acumulada con la frecuencia absoluta y realizar una interpretación confundiendo la variable esperanza de vida con número de personas.

5. Conclusiones y propuestas al profesorado en formación

A partir de los resultados encontrados, vemos que gran parte de las respuestas evidenciadas por este grupo de futuros profesores de Matemática están en un nivel inicial de compresión gráfica, es decir, establecen relaciones directas entre unas variables y algún elemento distintivo de un gráfico estadístico.

De los tres gráficos presentados en la actividad, la mayor cantidad de respuestas que mostraron un nivel inicial con interpretaciones correctas fueron las relacionadas con el gráfico de cajón y el histograma que, si bien se muestran en un nivel de lectura literal de los datos, son adecuadas, en comparación con la distribución de respuestas que se relevaron sobre la ojiva de manera incorrecta. En muy pocos casos los estudiantes evidencian un nivel de lectura más allá de los datos. Lo anterior es, por lo menos, cuestionable, considerando que los estudiantes para profesor que participaron de la investigación ya había cursado Estadística Descriptiva, Probabilidades e Inferencia Estadística.

En el caso de la interpretación de la ojiva los resultados son algo más preocupantes, dado que un alto porcentaje de profesores en formación no interpreta este gráfico, mostrando en sus respuestas aspectos de utilidad y descripción sobre la forma de uso. En muy pocos casos muestran un nivel inicial (lectura literal) y, entre quienes lo hacen también, una cantidad no menor presenta errores en la interpretación.

De estos resultados, y los niveles de lectura evidenciados, vemos que gran parte de los futuros profesores presenta dificultades en las interpretaciones de algunos estadísticos, o los interpreta de manera limitada y descontextualizada, aspectos que están en la misma línea de los trabajos realizados por Arteaga (2011) y Gea et al. (2017). Por ejemplo, confunden el significado de la variable esperanza de vida con la de promedio. Interpretan la media aritmética de manera descontextualizada sin hacer uso, por ejemplo, de la desviación estándar, y en algunos casos confunden e interpretan la mediana como si fuera la media aritmética.

Vemos con algo de preocupación que las interpretaciones en su mayoría están centradas en un nivel inicial de comprensión, considerando que los estudiantes, al momento de realizar el trabajo, cursaban su cuarto año de formación, lo que podría incidir en las dinámicas de enseñanza al momento de ejercer como profesores. En ninguna de las respuestas se visualizaron aspectos relativos a la variabilidad en relación con la variable esperanza de vida o una adecuada formalización de la asimetría de la distribución; no hay evidencias sobre un posible razonamiento inferencial, que permitiría ver una comprensión en un nivel más allá de los datos, como también resulta preocupante la gran cantidad de respuestas en relación con la interpretación que son incorrectas.

Tal como mencionamos, gran parte de estos resultados son preocupantes y no se pueden pasar por alto, pues si estas evidencias no son consideradas en el contexto de profesores de Matemática en formación, el impacto en la forma en que se enseña estadística seguirá la línea de lo que ya mencionan Cuétara et al. (2016), Del Pino y Estrella (2012), Pajares y Tomeo (2009), Zapata Cardona (2016) y Zapata Cardona y González Gómez (2017), es decir, la idea de estos aportes es poder potenciar la formación de los futuros profesores de Matemática que enseñarán estadística en contexto escolar para transitar a una enseñanza que profundice en la naturaleza del pensamiento estadístico y, por ende, en la cultura estadística de los futuros ciudadanos.

Referencias

Aoyama, K. (2007). Investigating a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education, 2*(3), 298-318. https://www.iejme.com/article/investigating-a-hierarchy-of-students-interpretations-of-graphs

Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores* [Tesis Doctoral. Universidad de Granada]. https://www.ugr.es/~batanero/pages/ARTICULOS/arteaga.pdf

Arteaga, P., Batanero, C., Cañadas, G., y Contreras, J. M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números. Revista de Didáctica de las Matemáticas, 76,* 55-67.

Batanero, C. (2002). *Los retos de la cultura estadística* [Conferencia inaugural]. Jornadas Interamericanas de Enseñanza de la Estadística, Buenos Aires, Argentina.

Bolch, C. A., y Jacobbe, T. (2019). Investigating Levels of Graphical Comprehension Using the LOCUS Assessments. *Numeracy, 12*(1), 158-174. https://doi.org/10.5038/1936-4660.12.1.8

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., y Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education, 49*(1), 2-8. https://doi.org/10.5951/jresematheduc.49.1.0002

Carmona, D., y Cruz, D. (2016). *Niveles de comprensión de la información contenida en tablas y gráficas estadísticas: un estudio desde la jerarquía de Kazuhiro Aoyama* [Tesis de Maestría, Universidad de Medellín]. Repositorio digital de documentos en Educación Matemática. http://funes.uniandes.edu.co/11422/1/Carmona2016Niveles.pdf

Centro de Estudios de Políticas y Prácticas en Educación. (2012). *Estándares orientadores para carreras de pedagogía en educación media*. CEPPE, Ministerio de Educación.

Cuétara, Y., Salcedo, I. M., y Hernández, M. (2016). La enseñanza de la estadística: antecedentes y actualidad en el contexto internacional y nacional. *Atenas, 3*(35), 125-140. http://atenas.umcc.cu/index.php/atenas/article/view/222/411

Curcio, F. R. (1989). *Developing graph comprehension.* National Council of Teachers of Mathematics.

Del Pino, G., y Estrella, S. (2012). Educación estadística: relaciones con la matemática. *Pensamiento Educativo, Revista De Investigación Latinoamericana (PEL), 49*(1), 53-64. https://doi.org/10.7764/PEL.49.1.2012

Díaz-Levicoy, D. (2018). *Comprensión de gráficos estadísticos por alumnos chilenos de Educación Primaria* [Tesis Doctoral, Universidad de Granada]. Repositorio institucional de la Universidad de Granada. https://digibug.ugr.es/bitstream/handle/10481/53598/29122880.pdf?sequence=4&isAllowed=y

Díaz-Levicoy, D., y Arteaga, P. (2019, febrero). *Chilean primary school diffculties in building bar graphs* [Conferencia]. Eleventh Congress of the European Society for Research in Mathematics Education (CERME11), Utrecht, Netherlands.

Fernández, N., García-García, J. I., Arredondo, E., y López, C. (2019). Comprensión de una tabla y un gráfico de barras por estudiantes universitarios. *Areté. Revista Digital del Doctorado en Educación de la Universidad Central de Venezuela, 5*(10), 145-162. http://saber.ucv.ve/ojs/index.php/ rev_arete/article/view/16992

Fitzallen, N. (2016). Interpreting association from graphical displays. En B. White y J. Clark (Eds.), *Opening Up Mathematics Education Research* (pp. 220-227). Mathematics Education Research Group of Australasia.

Friel, S., Curcio, F., y Bright, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in mathematics Education, 32*(2), 124-158. https://doi.org/10.2307/749671

Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review, 70*(1), 1-25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gal, I., y Murray, S. T. (2011). Responding to diversity in users' statistical literacy and information needs: Institutional and educational implications. *Statistical Journal of the International Association for Official Statistics, 27(*3-4), 185-195.

García-García, J. I., Encarnación Baltazar, E. J., y Arredondo, E. H. (2020). Exploración de la comprensión gráfica de estudiantes de secundaria. *IE Revista De Investigación Educativa De La REDIECH, 11,* e925. https://doi.org/10.33010/ie_rie_rediech.v11i0.925

Gea, M. M., Arteaga, P., y Cañadas, G. R. (2017). Interpretación de gráficos estadísticos por futuros profesores de Educación Secundaria. *Avances de Investigación en Educación Matemática, 12,* 19-37. https://doi.org/10.35763/aiem.v1i12.189

Hernández, R., Fernández, C., y Baptista, P. (2010). *Metodología de la investigación.* McGraw Hill.

Hüffmeier, J., Mazei, J., y Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology, 66*, 81-92. https://doi.org/10.1016/j.jesp.2015.09.009

Meletiou, M., y Lee C. (2002, 7 al 12 de julio). *Student understanding of histograms: A stumbling stone to the development of intuitions about variation* [Conferencia]. Sixth International Conference on Teaching Statistics, Durban, South Africa.

Meletiou-Mavrotheris, M., y Stylianou, D. A. (2003). Graphical representation of data: the effect of the use of dynamical statistics technological tool. En *CBLIS Conference Proceedings 2003 Volume I: New Technologies and their applications in education* (pp. 296-306). Department of Educational Sciences, University of Cyprus.

Ministerio de Educación, Cultura y Deporte de España. (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.* Autor.

Ministerio de Educación de Chile. (2015). *Bases curriculares 7° básico a 2° Medio.* Unidad de Currículum y Evaluación.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Autor.

Pajares, A., y Tomeo, V. (2009). Enseñanza de la Estadística y la Probabilidad en Secundaria: experimentos y materiales. En M. J. González, M. T. González y J. Murillo (Eds.), *Investigación en Educación Matemática. Comunicaciones de los grupos de investigación. XIII Simposio de la SEIEM.* Sociedad Española de Investigación en Educación Matemática.

Pfannkuch, M. (2006). Informal inferential reasoning. En A. Rossman y B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* International Statistical Institute and International Association for Statistical. http://www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf

Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education, 2*(3), 149-167. http://www.iejme.com/

Porta, L., y Silva, M. (2003). *La investigación cualitativa: el análisis de contenido en la investigación educativa.* CENIDE.

Sánchez, M. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education, 18*, 37-50. https://doi.org/10.1007/s10763-020-10069-7

Shaughnessy, J. M., Garfield, J. y Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (v.1, pp. 205-237). Dordrecht, Netherlands: Kluwer.

Schield, M. (2006). Statistical literacy survey analysis: Reading graphs and tables of rates and percentages. En A. Rossman y B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (pp. 1-6). International Statistical Institute and International Association for Statistical Education.

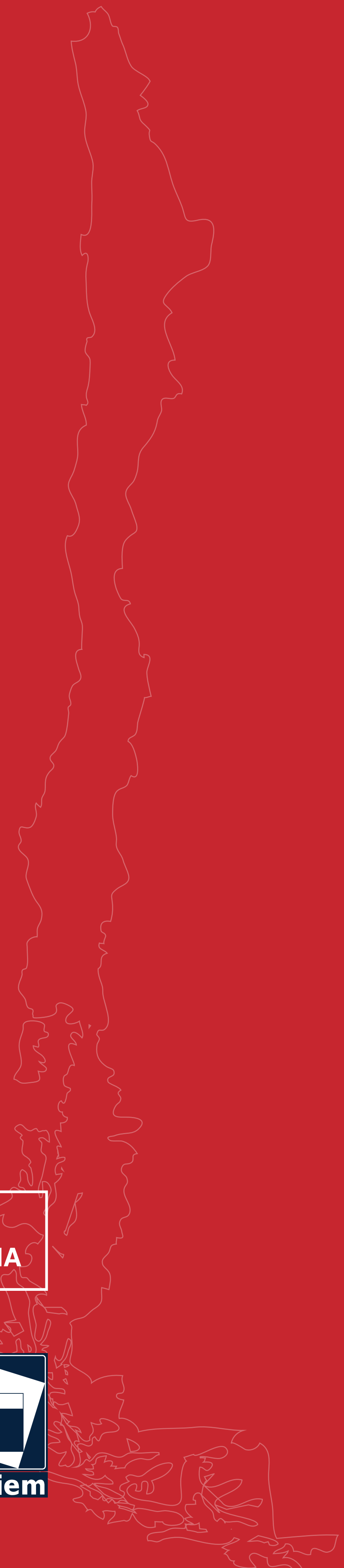
Vasilachis, I. (2006). *Estrategias de investigación cualitativa*. Gedisa.

Vásquez, C. (2021). Comprensión y Uso Docente de Gráficos Estadísticos por Futuros Profesores para Promover Competencias para la Sostenibilidad. *PARADIGMA, 41*(e1), 165-190. https://doi.org/10.37618/PARADIGMA.1011-2251.2021.p165-190.id1022

Zapata Cardona, L. (2016). ¿Estamos promoviendo el Pensamiento Estadístico en la enseñanza? En I. Álvarez y C. Sua (Eds.), *Memorias del II Encuentro Colombiano de Educación Estocástica* (pp. 73-79). Asociación Colombiana de Educación Estocástica.

Zapata Cardona, L., y González Gómez, D. (2017). Imágenes de los profesores sobre la estadística y su enseñanza. *Revista Educación Matemática, 29*(1), 61-89. https://doi.org/10.24844/EM2901.03

Zapico, M. (2007). Interrogantes acerca del análisis de contenido y del discurso en los textos escolares. En *Primer Seminario Internacional de textos Escolares (SITE 2006)* (pp. 149-155). Ministerio de Educación de Chile.



R E C H I E M

REVISTA CHILENA DE EDUCACIÓN MATEMÁTICA

V.13
NÚMERO ESPECIAL
- 4 -

LA EDUCACIÓN ESTADÍSTICA EN EL AULA ESCOLAR CHILENA

sochiem