

Conocimiento estadístico del futuro profesorado de educación primaria en la representación de datos

Pre-service primary school teachers' statistical knowledge in data representation

Ane Izagirre @ , Jon Anasagasti @ , Ainhoa Berciano @ 

Universidad del País Vasco/Euskal Herriko Unibertsitatea (España)

Resumen ∞ Este artículo analiza si el futuro profesorado de educación primaria distingue las categorías/valores de la variable de las frecuencias absolutas, selecciona y construye el gráfico adecuado según el tipo de variable y cómo justifica tal elección. Así, presentamos un estudio de caso de corte mixto con predominancia cualitativa y carácter exploratorio, con una muestra de 108 futuros docentes en el que se ha diseñado un cuestionario ad-hoc con 4 ítems. Los resultados indican que el futuro profesorado tiene dificultades para distinguir las categorías de una variable cualitativa de sus frecuencias absolutas. Los errores más comunes en la elección del gráfico son realizar gráficos de líneas para variables cualitativas e histogramas para variables cualitativas o cuantitativas discretas sin agrupar. Tanto el grado de corrección en la realización del gráfico como su justificación no es suficiente, lo que evidencia la necesidad de mejorar su alfabetización estadística.

Palabras clave ∞ Alfabetización estadística; Futuros docentes de primaria; Medidas centrales; Gráficos estadísticos; Justificación estadística

Abstract ∞ This article analyses if pre-service primary school teachers distinguish the categories or values of the statistical variable from the absolute frequencies, whether they select and construct the appropriate graph according to the type of variable and how they justify such choice. Thus, we present a mixed-method case study with qualitative predominance and exploratory nature. The sample is conformed of 108 future primary education teachers, and the questionnaire is an ad-hoc questionnaire with 4 items. The results indicate that pre-service teachers have difficulties in distinguishing the categories of a qualitative variable from its absolute frequencies. The most common error detected in the graph's choice is the selection of the line graph for qualitative variables and histograms for qualitative or non-grouped discrete quantitative variables. The correction degree in the realization of the graph and its justification are not sufficient. This shows the need to improve the statistical literacy of pre-service teachers.

Keywords ∞ Statistical literacy; Pre-service primary school teachers; Central tendency measures; Statistical graphs; Statistical reasoning

Izagirre, A., Anasagasti, J. & Berciano, A. (2023). Conocimiento estadístico del futuro profesorado de educación primaria en la representación de datos. *AIEM - Avances de investigación en educación matemática*, 24, 111-130. <https://doi.org/10.35763/aiem24.4646>

1. INTRODUCCIÓN

En la sociedad en la que vivimos, para llegar a formar una ciudadanía bien informada y consumidora crítica con su entorno, ésta debe saber analizar datos y tener ciertas nociones probabilísticas para razonar estadísticamente, lo que conlleva a que el alumnado deba ser instruido en este ámbito (National Council of Teachers of Mathematics, 2003). Batanero y Godino (2002) señalan que la enseñanza de la estadística fomenta el razonamiento crítico y que es de gran utilidad para comprender los contenidos de otras asignaturas, durante la etapa académica, y para su posterior aplicación en el ámbito laboral. Por todo ello, es importante trabajar el tratamiento de la información y la probabilidad desde educación infantil (Alsina, 2012). Para asegurar que todo el alumnado recibe una formación matemática de calidad en estadística y probabilidad es necesario, entre otras cosas, que el futuro profesorado tenga un conocimiento disciplinar profundo en estos dos ámbitos.

Sin embargo, las investigaciones hasta ahora realizadas muestran una realidad muy distinta. Si hacemos un breve repaso de los estudios llevados a cabo con futuro profesorado encontramos que, para el caso de educación infantil, Berciano et al. (2021) ponen de manifiesto la poca importancia que este colectivo da a la estadística en contextos STEAM y la clara confusión que tienen sobre qué son los gráficos estadísticos. Igualmente, para el caso de educación primaria, Estrada (2007) afirma que, en general, el futuro profesorado comete errores conceptuales en algunos de los conocimientos estadísticos elementales. Uno de los motivos de estos errores es, según Anasagasti (2019), que hay estudiantes que apenas han recibido instrucción estadística previa. Así, con el propósito de ofrecer al futuro profesorado una formación ajustada a sus necesidades, en este trabajo planteamos una investigación que ahonde en detectar cuáles son los errores más frecuentes que comete el profesorado en formación inicial cuando se le plantean tareas básicas de tratamiento e interpretación de información por medio de estadísticos centrales y gráficos estadísticos. Así, los tres objetivos específicos que planteamos son: 1) investigar si el futuro docente de educación primaria diferencia las categorías o valores de la variable estadística de las frecuencias absolutas en el cálculo de la moda; 2) analizar si el futuro docente de primaria hace una selección adecuada del tipo de gráfico según el tipo de datos proporcionados, esto es, según el tipo de variable; y 3) evaluar la corrección de los gráficos construidos por el futuro docente de primaria y analizar la justificación que da para su construcción.

2. MARCO TEÓRICO Y ANTECEDENTES

Con el fin de que la ciudadanía sea crítica con su entorno y pueda tomar decisiones con respecto a la información referente a datos y probabilidades que nos rodean, es necesario que ésta tenga desarrollado, con garantías, el sentido estocástico.

Esta necesidad ha sido puesta de manifiesto en los currículos educativos de distintos países en los últimos años. En el caso de España, los estándares de aprendizaje evaluables que se recogen en el Currículo de educación primaria definido por la LOMCE (Real Decreto 126/2014, p. 38) ya señalaban que en el bloque relativo a “Estadística y Probabilidad” el alumnado de primaria debe:

Identificar datos cualitativos y cuantitativos; recoger y clasificar datos cualitativos y cuantitativos utilizándolos para construir tablas de frecuencias absolutas y relativas; aplicar de forma intuitiva las medidas de centralización: la media aritmética y la moda; y realizar e interpretar gráficos muy sencillos: diagramas de barras, poligonales y sectoriales.

Igualmente, la LOMLOE, que está ya en vigor y se implantará paulatinamente, añade la importancia de trabajar explícitamente el *sentido estocástico* con el fin de desarrollar la competencia matemática, entendido este como “el razonamiento y la interpretación de datos y la valoración crítica, así como la toma de decisiones a partir de información estadística” (Real Decreto 157/2022, p. 101). Dicho sentido recoge los saberes mínimos en el Bloque E, entre los que destacamos “Representación de datos obtenidos a través de recuentos mediante gráficos estadísticos sencillos” y “La moda: interpretación como el dato más frecuente” (Real Decreto 157/2022, pp. 108-112).

Con el fin de analizar las dificultades que surgen en la adquisición de las competencias asociadas a este bloque de contenidos se han realizado distintos estudios en diferentes etapas educativas que ponen de manifiesto que algunos errores, una vez adquiridos, se mantienen a lo largo de toda la educación formal. Con respecto a estas investigaciones, la mayoría se centra en el tercer y cuarto estándar, apareciendo el primero y el segundo de forma indirecta, pero no por ello son menos importantes.

Con respecto al tercer estándar, las medidas de centralización, son varias las investigaciones que detallan las dificultades de estudiantes de educación primaria y secundaria. Así, Cobo (2003), Batanero et al. (2003), Mayén et al. (2007) y Carvalho (2001) detectan errores en el cálculo de la media y mediana. Para el caso de la media: descuidar los valores cero de la variable, calcular el promedio de los valores que toma la variable olvidándose de las frecuencias absolutas; calcular el promedio de las frecuencias absolutas, entre otros. Para el cálculo de la mediana: no ordenar los datos de forma creciente; para un tamaño par de la muestra, no saber resolver dicha indeterminación; o cuando los datos están agrupados, tener en cuenta solamente los valores, descuidando las frecuencias absolutas. En el caso del alumnado universitario, otro error detectado por Madrid et al. (2022) es ordenar los valores de la variable en función de su frecuencia, en lugar de utilizar el orden de sus valores.

Restringiendo nuestro interés al caso de la moda, las investigaciones que abordan los errores en su cálculo son escasas. Entre ellas, Carvalho (2001) encuentra que el alumnado de secundaria da este dato como respuesta al cálculo de la mediana, o expresa como respuesta la mayor frecuencia absoluta; igualmente, Freitas et al. (2018) ponen de manifiesto que el alumnado confunde la mediana con la media o la moda. Cuando la población de estudio es el profesorado de educación primaria (en formación o en activo), tal como indica el estudio de Sanoja de Ramírez y Ortiz-Buitrago (2013), la moda se suele excluir de las investigaciones que describen las dificultades para el cálculo de medidas de tendencia central, debido a que todo el profesorado cree conocer cómo se determina. Sin embargo, los pocos estudios previos existentes revelan que el profesorado sí tiene dificultades con esta medida

central. Leavy y O'loughlin (2006) concluyen que el 25 % de los futuros docentes confunde la media con la moda. Estrada et al. (2004) llegan al mismo resultado en una investigación sobre los conceptos estadísticos elementales, observando que los futuros docentes confunden la media, mediana y moda en distribuciones asimétricas. Groth y Bergner (2006) concluyen que estos relacionan el cálculo de la moda únicamente con las variables cuantitativas.

Sobre el cuarto estándar, los estudios realizados con alumnado de educación primaria y secundaria sobre la interpretación de gráficos revelan que éste tiene dificultades en la interpretación de gráficos estadísticos, asociando estas dificultades y niveles de lectura al tipo de gráfico seleccionado (Arteaga et al., 2021; Fernandes y Morais, 2011). Igualmente, Friel y Bright (1996) mencionan que el alumnado confunde los valores de la variable de las frecuencias absolutas cuando los datos se muestran mediante un gráfico.

En cuanto a la representación de datos mediante gráficos estadísticos, estudios previos determinan que los errores provienen, primeramente, de la elección, y seguidamente, de la corrección del gráfico creado.

Sobre la elección del gráfico estadístico, para el caso de educación primaria y secundaria, uno de los resultados más notables es que el alumnado hace una selección incorrecta del gráfico para el tipo de variable que se le presenta (Li y Shen, 1992; Wu, 2004, p. 37; citado en Díaz-Levicoy, 2018). Esta confusión no es ajena para el caso de futuros docentes de educación primaria, encontrando investigaciones que determinan la elección del histograma para variables cuantitativas discretas sin agrupar (Arteaga et al., 2016; Bruno y Espinel, 2005) o el uso indistinto del diagrama de barras y el histograma (Espinel, 2007); concluyendo que el problema está en que los participantes no distinguen lo cualitativo, lo discreto y lo continuo. Sin embargo, en una investigación donde los futuros docentes debían elegir entre 4 gráficos previamente dados para representar un conjunto de datos (gráfico de barras, sectores, líneas y diagrama de dispersión) y justificar dicha elección, Alacaci et al. (2011) señalan que el futuro profesorado tiene un conocimiento profundo sobre el gráfico de líneas relacionándolo con situaciones que muestran una tendencia.

Para el caso de futuros docentes de educación primaria, Arteaga y Batanero (2010) y Batanero et al. (2009) identifican los siguientes errores en la construcción de gráficos: a) altura de la barra, punto o ángulo no proporcional a la frecuencia; b) intercambio de frecuencia y valor de la variable en los ejes; c) representación del valor de la variable junto con su frecuencia; y d) representación de la variable multiplicado por su frecuencia. Y Arteaga (2011) detecta los siguientes errores: a) unión de puntos del gráfico de líneas con líneas curvas; b) no centrar las barras del histograma en la marca de clase; c) omitir valores de frecuencia nula en los gráficos de barras, polígonos e histogramas; d) elección de escalas demasiado amplias o que no cubren el recorrido de la variable; y e) subdivisión incorrecta de la escala.

3. MÉTODO

El trabajo realizado se ha llevado a cabo desde un paradigma interpretativo y se enmarca dentro de las investigaciones de corte mixto, con una predominancia

cualitativa y carácter exploratorio, clasificando e interpretando tanto la producción de gráficos estadísticos como la justificación de su elección.

3.1. Muestra

La muestra está compuesta por 108 estudiantes de tercer curso del Grado de Educación Primaria de la Universidad del País Vasco. Todos ellos han cursado la asignatura *Matemáticas y su Didáctica II*, asignatura anual de 9 ECTS que abarca cuatro bloques curriculares: Números y operaciones, geometría, la medida, y estadística, y probabilidad. Al bloque de la estadística y probabilidad se destinan 24 horas que es todo el contenido del bloque curricular que reciben a lo largo del Grado. Los datos se tomaron una vez cursado este bloque de contenidos. La realización del cuestionario fue voluntaria y se ha respetado la anonimidad de las y los participantes.

3.2. Instrumento de recogida de información

El instrumento es un cuestionario que contiene 4 ítems (Tabla 1). En el primero la variable es cualitativa y en el resto las variables son cuantitativas discretas. Se han omitido las variables continuas porque, como indica Calot (1974; citado en Cobo 2003), la moda no está bien definida si la amplitud de los intervalos de clase es diferente, a no ser que se aplique una fórmula (Cobo, 2003, p. 56), lo cual dificultaría el cálculo y cabría la posibilidad de cometer errores procedimentales que no son objeto de estudio. La elección de las preguntas del cuestionario se basa en que, para responder correctamente, es necesario distinguir debidamente los valores o las categorías que toma la variable de las frecuencias absolutas.

Cabe señalar que los y las participantes han realizado el cuestionario en Euskara. Por lo tanto, las respuestas que se muestran más adelante también están en Euskara. Los participantes han requerido entre 17-35 minutos para responder el cuestionario.

Tabla 1. Ítems y objeto de estudio de cada ítem

Ítem	Objeto de estudio
<p>1. Un vendedor de globos tiene los siguientes globos. Basándote en el color de los globos:</p> <p>(i) Construye la tabla de frecuencias.</p> <p>(ii) ¿Cuál es la moda?</p> <p>(iii) Representa los datos mediante una gráfica y justifica dicha elección.</p>	<ul style="list-style-type: none"> - Distinción entre las categorías de la variable cualitativa “color” (verde, rojo, amarillo y azul) y las correspondientes frecuencias absolutas al extraer la información de una imagen. - Tabla de frecuencias. A pesar de que hay 11 globos, algunos participantes solamente han visto 10, des-cuidando un globo rojo escondido. Se han considerado correctas ambas observaciones. - La moda. - Justificación de la elección del gráfico. - Construcción del gráfico.



Ítem	Objeto de estudio												
<p>2. Se ha preguntado a un grupo de jóvenes cuál es la paga semanal que reciben (€) y esto es lo que han contestado:</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> 5, 10, 15, 10, 5, 10, 15, 15, 10, 5, 10, 5, 15, 5, 5, 10, 15, 15, 15, 10, 10, 10, 15, 10, 15, 15, 15, 5, 10, 5, 10, 20, 15, 15, 5 </div> <p>(i) Construye la tabla de frecuencias. (ii) ¿Cuál es la moda? (iii) Representa los datos mediante una gráfica y justifica dicha elección.</p>	<ul style="list-style-type: none"> - Distinción entre los valores de la variable cuantitativa discreta “cantidad de euros” (5€, 10€, 15€ y 20€) y las correspondientes frecuencias absolutas (que se aproximan a los valores que toma la variable) al extraer la información desde una lista de valores discretos. - Tabla de frecuencias. - La moda. - Justificación de la elección del gráfico. - Construcción del gráfico. 												
<p>3. El diagrama de sectores recoge el número de caries que ha detectado un dentista a 100 niños y niñas en un colegio.</p> <p>(i) Construye la tabla de frecuencias. (ii) ¿Cuál es la moda?</p>	<ul style="list-style-type: none"> - Distinción entre los valores de la variable cuantitativa discreta “número de caries” (0, 1, 2, 3 y 4) y las correspondientes frecuencias relativas al extraer la información desde un diagrama de sectores. - Tabla de frecuencias. - La moda. 												
<p>Número de caries</p> <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <thead> <tr> <th>Número de caries</th> <th>Frecuencia relativa (%)</th> </tr> </thead> <tbody> <tr> <td>0 caries</td> <td>25 %</td> </tr> <tr> <td>1 caries</td> <td>20 %</td> </tr> <tr> <td>2 caries</td> <td>35 %</td> </tr> <tr> <td>3 caries</td> <td>15 %</td> </tr> <tr> <td>4 caries</td> <td>5 %</td> </tr> </tbody> </table>		Número de caries	Frecuencia relativa (%)	0 caries	25 %	1 caries	20 %	2 caries	35 %	3 caries	15 %	4 caries	5 %
Número de caries	Frecuencia relativa (%)												
0 caries	25 %												
1 caries	20 %												
2 caries	35 %												
3 caries	15 %												
4 caries	5 %												
<p>4. El número de errores realizados por el alumnado de la asignatura Matemática y su didáctica II es:</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> 1, 2, 1, 3, 4, 2, 2, 1, 1, 3, 3, 1, 2, 3, 4, 4, 1, 2, 3, 3, 2, 2, 1, 1, 13, 4, 1, 2, 4, 3, 2, 1, 2, 3, 1, 1, 1, 2, 2, 1, 2, 2, 2, 1, 3, 3, 1, 1 </div> <p>(i) Construye la tabla de frecuencias. (ii) ¿Cuál es la moda? (iii) Representa los datos mediante una gráfica y justifica dicha elección.</p>	<ul style="list-style-type: none"> - Distinción entre los valores de la variable cuantitativa discreta “número de errores” (1, 2, 3, y 4) y las correspondientes frecuencias absolutas (que difieren de los valores que toma la variable) al extraer la información desde una lista de valores discretos. - Tabla de frecuencias. - La moda. - Justificación de la elección del gráfico. - Construcción del gráfico. 												

3.3. Rúbricas para la categorización del grado de corrección del gráfico realizado y de la tipología de justificación dada

Se presentan las rúbricas mediante las cuales se han clasificado las respuestas del alumnado de los ítems 1, 2 y 4 según el nivel de corrección del gráfico (Tabla 2) y el tipo de justificación (Tabla 3).

Para la rúbrica sobre niveles de corrección de los gráficos estadísticos (Tabla 2) se han tenido en cuenta estudios previos que clasifican gráficos por su corrección (Arteaga, 2011) y se han desglosado algunas de las categorías descritas en dicho trabajo para analizar en mayor profundidad los resultados de esta investigación. Considerando que los gráficos estadísticos se han realizado a mano, la mayoría con bolígrafo sobre papel, y que no pueden ser absolutamente precisos, se han definido cinco niveles de corrección.

Tabla 2. Niveles de corrección de los gráficos estadísticos realizados

Nivel	Descripción	Arteaga (2011)
CG1	Realizado correctamente: el gráfico representa correctamente la distribución de los datos. Si es un diagrama de barras o histograma, los ejes están bien contruidos y la escala es visible y correcta; si son diagramas de sectores, se indican las categorías y los porcentajes o las frecuencias.	E1. Gráficos básicamente correctos: gráficos correctos (título, ejes, escalas, etiquetas, tipo de gráfico y las variables representadas), básicamente correctos (pero no estándar) y gráficos con líneas innecesarias que dificultan su lectura.
CG2	Errores de rotulación: el gráfico representa correctamente la distribución de los datos en cada caso (las barras y los sectores son proporcionales), pero hay elementos del gráfico que no se indican (rótulos de los ejes, categorías/valores, escalas o frecuencias absolutas/relativas).	E2. Gráficos parcialmente correctos con errores en escala (escalas no apropiadas o proporcionales, rótulos confusos, ...).
CG3	Errores de proporcionalidad: falta de correspondencia entre los pesos relativos de los datos y su representación geométrica.	
CG4	Otros errores no relacionados con la proporcionalidad: categorías identificadas incorrectamente, frecuencias mal representadas, intercambiar frecuencia por valor/categoría de la variable, ...	E3. Gráficos incorrectos (altura de la barra, punto, ángulo no proporcional a la frecuencia, intercambia frecuencia y valor de la variable...).
CG5	Elección incorrecta del gráfico.	

Por otro lado, la rúbrica mediante la cual se analiza la justificación que dan para su construcción (Tabla 3) es una adaptación de Berciano et al. (2021). Dichos niveles se corresponden con los cuatro niveles jerarquizados según el nivel de razonamiento, definidos en el estudio de Alacaci et al. (2011), si bien para el último nivel presentado en dicho estudio nuestra escala diferencia las justificaciones incorrectas de base estadística de las que no.

Tabla 3. Niveles de justificación para la selección del gráfico

Nivel	Descripción	Berciano et al. (2021)	Alacaci et al. (2011)
J1	Justifica correctamente: se argumenta la selección del gráfico en función del tipo de variable utilizando los términos correctos: cualitativa o cuantitativa (discreta/continua).	Nivel 4. Describe las cantidades de cada categoría/valor, estableciendo su peso relativo respecto al total.	A4. Explicación conceptual directa: la explicación de la elección (o no) del gráfico se refiere explícitamente al tipo de tarea implícita en el escenario.
J2	Justifica y utiliza los términos correctamente, pero usa la palabra "variable" al referirse a categorías o valores determinados.	Nivel 3. Describe las cantidades mencionando las categorías/valores.	A3. Explicación conceptual indirecta: la explicación de la elección (o no) del gráfico hace referencia indirectamente al tipo de tarea.
J3	A pesar de no nombrar el tipo de variable, expresa características de ellas. Implícitamente se puede llegar a entender qué quiere decir.	Nivel 2. Describe las cantidades sin mencionar las categorías/valores.	A2. Explicación sobre los componentes estructurales de los gráficos: la explicación considera los componentes estructurales de los gráficos (barras, puntos, líneas, etiquetas, etc.).
J4	Utiliza expresiones incorrectas para justificar estadísticamente su elección.	Nivel 1. Menciona las categorías/valores.	
J5	Incorrectos: justificaciones que no tienen base estadística.	Nivel 0. No da ninguna argumentación estadística.	A1 Otras explicaciones: la explicación se basa en otras razones como opiniones o preferencias personales.

3.4. Procedimiento para la categorización del grado de corrección del gráfico realizado y de la tipología de justificación dada

Para interpretar la información recogida en los cuestionarios y aplicar las rúbricas, se ha realizado un proceso de triangulación entre investigadores: inicialmente, cada investigador individualmente ha clasificado las respuestas de toda la muestra del estudio; se han contrastado los resultados de los tres investigadores obteniendo un elevado índice de concordancia; finalmente, en el reducido porcentaje de discordancia entre los expertos, se han analizado grupalmente dichos casos quedando clasificados de manera consensuada.

4. RESULTADOS

Mostramos los resultados del estudio atendiendo a los tres objetivos específicos de investigación. Para ello, en cada apartado presentamos los resultados de forma global cuantitativamente, y, posteriormente, mostramos el análisis cualitativo de algunos de los casos más significativos.

4.1. Distinción entre las categorías o valores de la variable estadística y las frecuencias absolutas en el cálculo de la Moda

Para la variable cualitativa (ítem 1), dentro del conjunto de respuestas incorrectas (15.7 %) encontramos que las frecuencias absolutas han sido interpretadas como los valores de una nueva variable cuantitativa, asignándole, así, un valor cuantitativo a una variable cualitativa. El 10.2 % (11 pers.) ha identificado como moda la referida a la moda de las frecuencias absolutas (ver Figura 1), el 4.6 % (5 pers.) ha calculado la media aritmética de los valores máximos de las frecuencias absolutas (ver Figura 2) y el 0.9 % (1 pers.) ha calculado la mediana de las frecuencias absolutas (ver Figura 3).

Figura 1. Moda de las frecuencias absolutas para N=11 y N=10 globos, respectivamente

Zein da Moda? 3 Zein da Moda? 2,3 globo

Figura 2. Media aritmética de los valores máximos de las frecuencias absolutas para N=11 y N=10 globos, respectivamente

Zein da Moda? $\frac{3+3+3}{3} = 3$ Zein da Moda? x_1 (horria) eta x_3 (Berdea) $\rightarrow \frac{3+3}{2} = 3$ moda

Figura 3. Mediana de las frecuencias absolutas para N=10

Zein da Moda? 2,5 globo 2 2 3 3
2,5

Al analizar los resultados para variables cuantitativas, vemos que hay un porcentaje de participantes que realiza incorrectamente la tarea. En el ítem 2 son cinco las respuestas incorrectas (4.6 %): $M_o=5$, $M_o=10$, $M_o=10$, $M_o=10$ y 15 , $M_o=20$; pero en todos los casos las respuestas son valores de la variable. Un análisis más pormenorizado, nos lleva a determinar que en dos casos la tabla de frecuencias es incorrecta y en los otros tres casos parece haber un error procedimental a la hora de identificar la máxima frecuencia absoluta y el correspondiente valor de la variable. Además, hay un caso que responde “la moda es 13, es decir, salir de casa con 15€”. Se supone que el/la participante sabe definir la moda, aunque presenta dificultades para expresarse utilizando un vocabulario estadístico apropiado.

En el ítem 3, a pesar de que se puede observar gráficamente cuál es la moda, tres personas responden incorrectamente (2.7 %) y las soluciones dadas son: $M_o=0$, $M_o=4$ y $M_o=35$ (el valor de la frecuencia absoluta). Entendemos que las dos primeras respuestas son un error procedimental y no conceptual, ya que, 0 y 4 son valores de la variable. En cambio, la persona que responde $M_o=35$ no ha realizado la conexión entre la máxima frecuencia relativa y el valor de la variable, considerándolo, así, un error conceptual. Esta vez también la misma persona mencionada anteriormente responde “la moda es 35, es decir, tener 2 caries”.

En el ítem 4 son seis las respuestas incorrectas (5.5 %); una persona no relaciona la máxima frecuencia absoluta con el valor de la variable, dando como respuesta $M_o=18$ (el valor de la frecuencia absoluta), y cinco personas responden que la moda es 2. Entendemos que este último ha sido un error procedimental y no conceptual. Igualmente, la misma persona mencionada previamente da como respuesta “la moda es 18, es decir, hacer un error”.

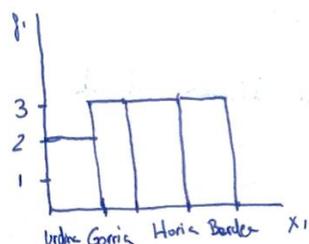
Contrastando los resultados descritos en este apartado, el grado de corrección en el cálculo de la moda es de 84.3 % para la variable cualitativa, frente a un 95.4 %, 97.2 % y 94.4 % para las variables cuantitativas de los ítems 2, 3 y 4, respectivamente.

4.2. Valoración global del tipo de gráfico según el tipo de variable

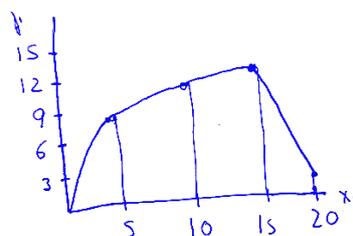
Un análisis global de los datos muestra que el 71.3 % de las y los participantes ha seleccionado correctamente todos los gráficos, el 13.88 % dos de los tres gráficos, el 10.19 % uno de los tres gráficos y el 3.7 % ningún gráfico. Entre las personas que han seleccionado correctamente dos gráficos, el 60 % ha elegido adecuadamente el gráfico correspondiente a la variable cualitativa (ítem 1) y una de las dos variables cuantitativas; y el 40 % los dos gráficos correspondientes a las variables cuantitativas (ítem 2 y 4). Asimismo, entre las personas que han seccionado correctamente un único gráfico, el 54.54 % lo ha hecho para la variable cualitativa (ítem 1) y el 45.45 % para una de las variables cuantitativas (ítem 2 o 4).

La Figura 4 muestra dos cuestionarios incorrectos donde todos los gráficos seleccionados son inadecuados para el tipo de variable de cada ítem.

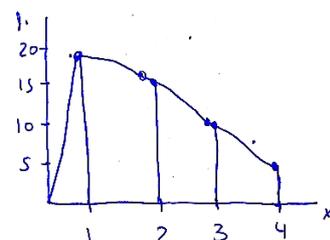
Figura 4. Elecciones incorrectas de los participantes D3112 y D3234 para los ítems 1, 2 y 4, respectivamente



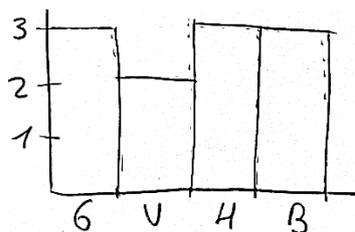
D3112. Ítem 1. Histograma.



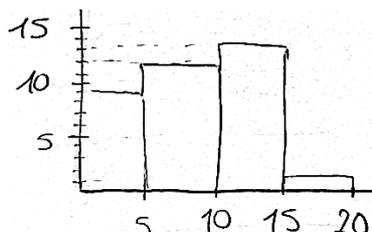
D3112. Ítem 2. Polígono de frecuencias.



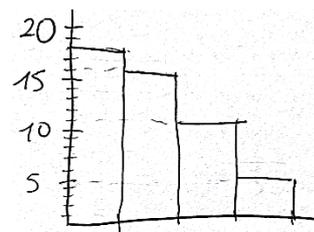
D3112. Ítem 4. Polígono de frecuencias.



D3234. Ítem 1. Histograma.



D3234. Ítem 2. Histograma.



D3234. Ítem 4. Histograma

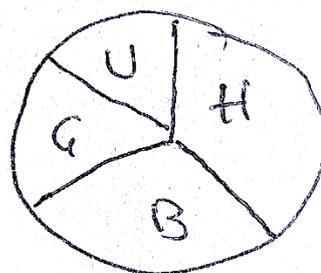
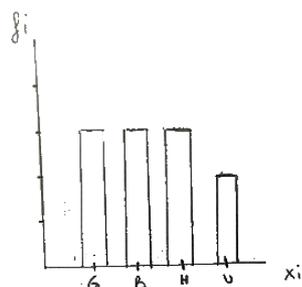
4.3. Clasificación de errores en la construcción de gráficos estadísticos

A continuación, mostraremos de forma cualitativa las representaciones más significativas intercalando el porcentaje de participantes en cada categoría.

Cuando la variable es cualitativa (ítem 1), a pesar de que el 82.42 % de los casos selecciona el gráfico apropiadamente, llama la atención la falta de corrección en su construcción. Un análisis más pormenorizado de los gráficos construidos nos lleva a identificar ausencias importantes:

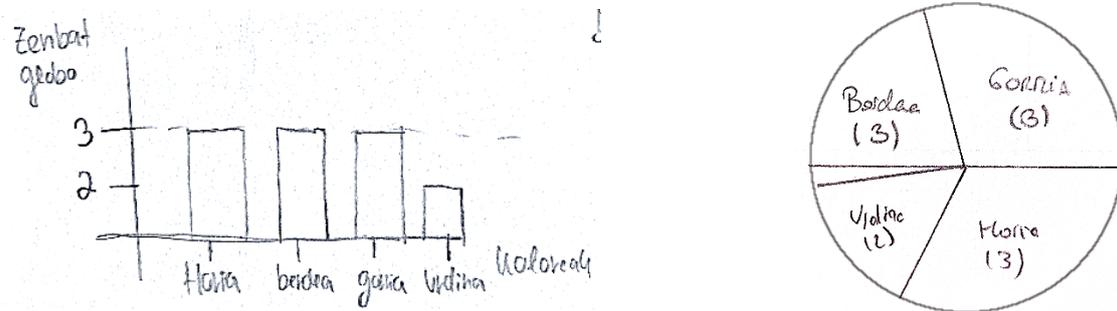
- Omisión de las categorías y/o frecuencias absolutas en los diagramas de barras, y las frecuencias absolutas o relativas en los diagramas de sectores (CG2; 25.93 %; ver Figura 5).

Figura 5. Representaciones de los participantes B3135 y D319, respectivamente, en el ítem 1 clasificadas en la categoría CG2



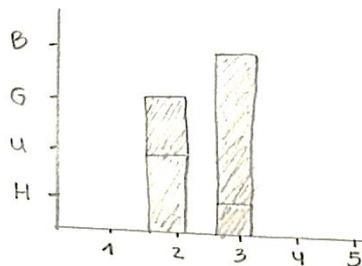
- Falta de proporcionalidad en los gráficos donde la escala de las frecuencias absolutas o los ángulos de los sectores no son proporcionales (CG3; 4.63 %; ver Figura 6).

Figura 6. Representaciones de los participantes B315 y D3118, respectivamente, en el ítem 1 clasificadas en la categoría CG3



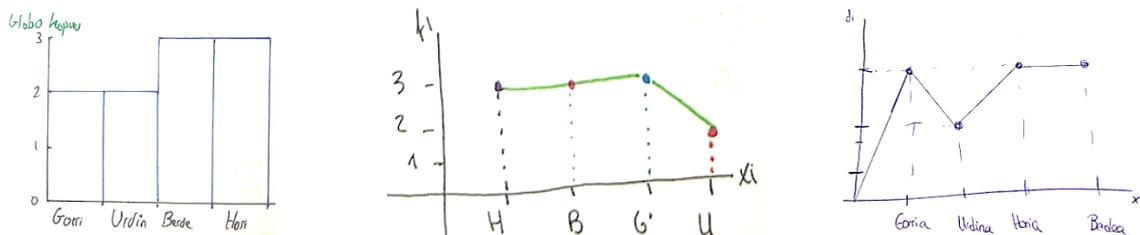
- Otros errores no relacionados con la proporcionalidad. Un participante ha representado los datos mediante un diagrama de barras apilado (CG4; 0.93 %; ver Figura 7).

Figura 7. Representación del participante D3241 en el ítem 1 clasificado en la categoría CG4



- Entre los gráficos mal seleccionados (CG5; 14.8 %), encontramos histogramas (8.3 %), gráfico de líneas (5.6 %) o polígonos de frecuencias (0.9 %), ver Figura 8.

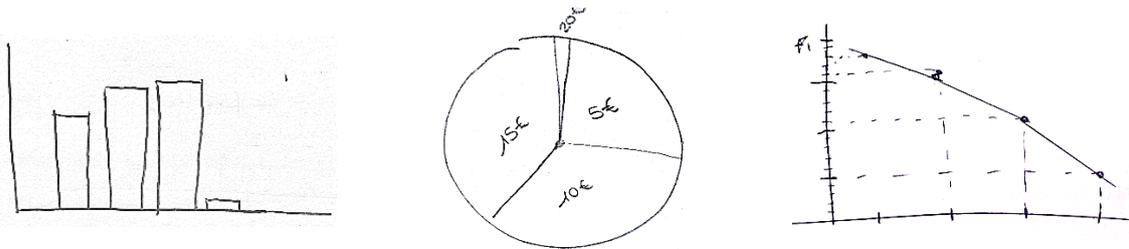
Figura 8. Elecciones incorrectas (CG5) en el ítem 1 de los participantes D3225, D318 y D317, respectivamente



Para el caso de una variable cuantitativa discreta, los ítems 2 y 4 muestran resultados similares. A pesar de que el 81.47 % y 85.2 % de los participantes, respectivamente, selecciona correctamente el gráfico, la corrección de los gráficos no es el esperado. Así, tanto en el ítem 2 como en el ítem 4, tenemos errores de construcción asociados a:

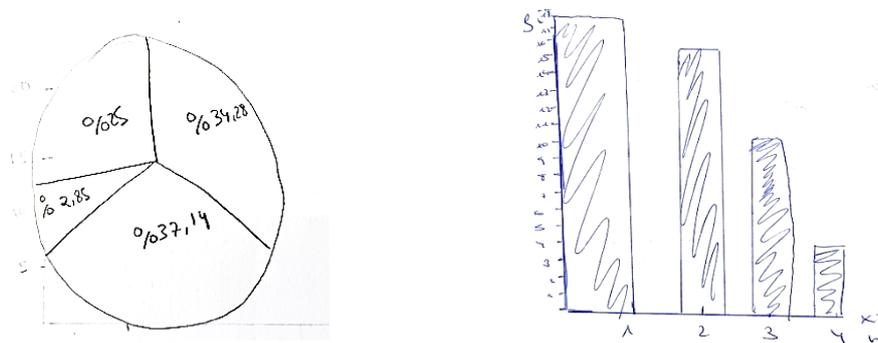
- No especificar los valores y/o frecuencias absolutas en los diagramas de barras y gráficos de línea, y las frecuencias absolutas o relativas en los diagramas de sectores (CG2; 28.7 % y 36.11 %, resp., ver Figura 9).

Figura 9. Representaciones de los participantes D3222, D3113 y D317 en los ítems 2 y 4, respectivamente, clasificadas en la categoría CG2



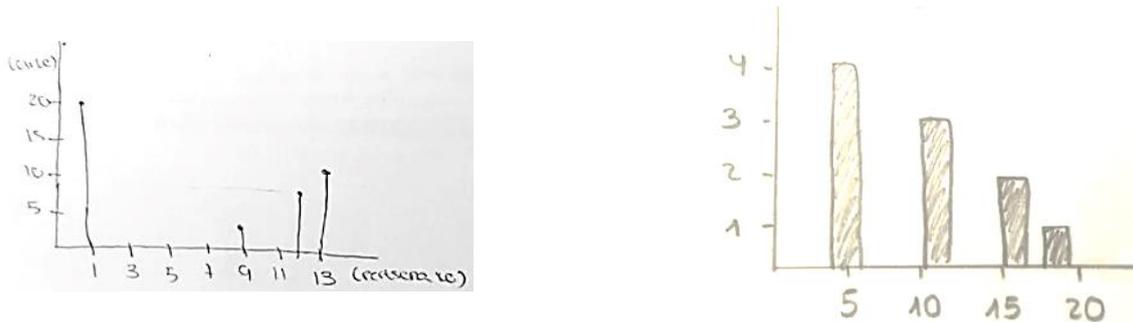
- Falta de proporcionalidad en los gráficos. En los diagramas de sectores los ángulos no son proporcionales y en los diagramas de barras la anchura de las barras no es proporcional (CG3; 1.85 % en ambos casos, ver Figura 10).

Figura 10. Representaciones de los participantes B3124 y B3110 en los ítems 2 y 4, respectivamente, clasificadas en la categoría CG3



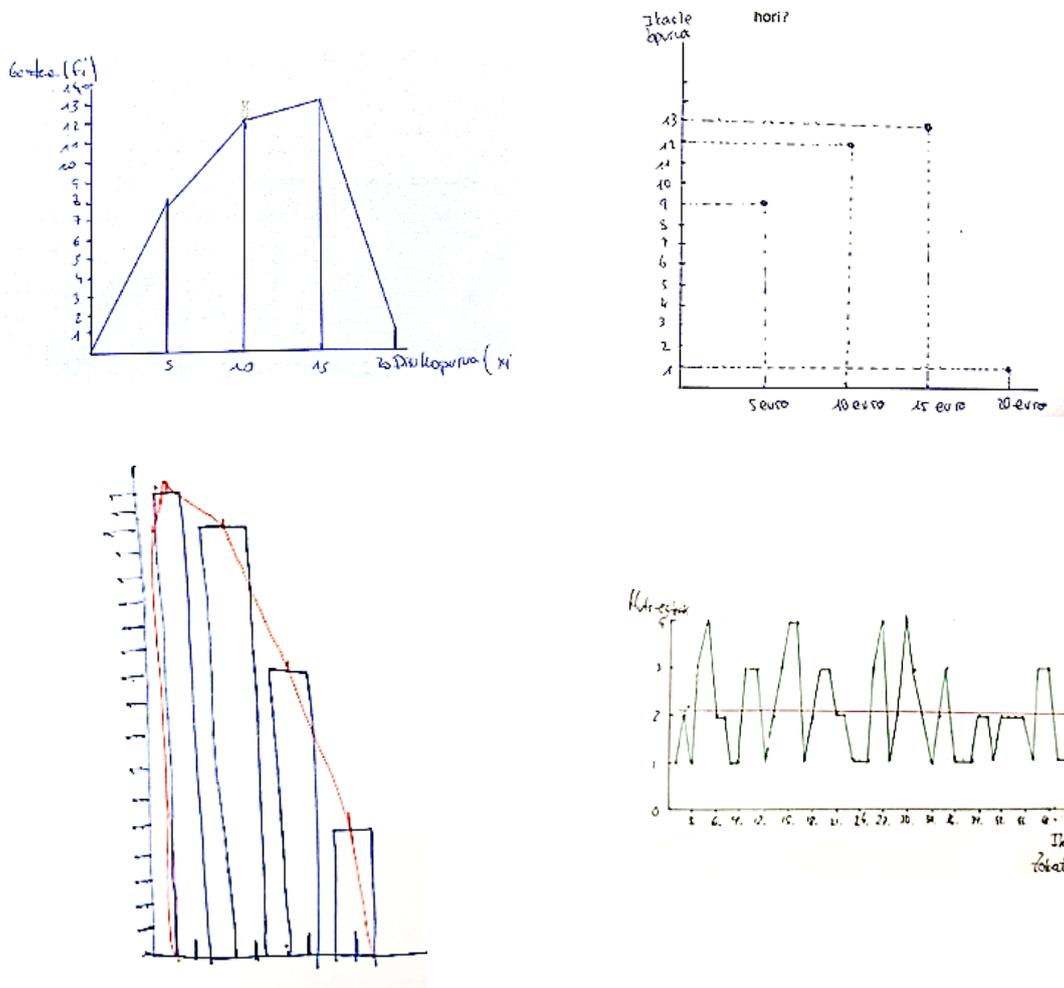
- Intercambiar los valores de la variable con las frecuencias absolutas (CG4; 1.85 % y 2.8 %, resp., ver Figura 11).

Figura 11. Representación de los participantes D3255 y D3241 en los ítems 2 y 4, respectivamente, clasificadas en la categoría CG4



- Los gráficos erróneos seleccionados (CG5) han sido, en su mayoría, histogramas (13.9 %, 11.1 %, resp.) y polígonos de frecuencias (3.7 %, 2.7 %, resp.), ver Figura 12.

Figura 12. Representaciones incorrectas (CG5) de los participantes D315, D3118, D3258 y D3225 en los ítems 2 y 4, respectivamente



4.4. Clasificación de las justificaciones utilizadas en la elección del gráfico estadístico

Con la intención de ver con mayor claridad las diferencias entre el alumnado a la hora de justificar correctamente la elección del gráfico estadístico, en la Figura 13 se muestra el nivel de justificación máximo alcanzado por el futuro profesorado de educación primaria. Solamente el 19 % del alumnado alcanza un nivel correcto de justificación (J1), haciendo un uso apropiado de los términos estadísticos. Entre el alumnado que justifica su selección mediante argumentos estadísticos, encontramos un 18 % que, a pesar de usar los términos correctos, confunde el concepto variable con el de categoría (J2); este alumnado hace referencia a que las variables (en lugar de referirse a las categorías) son de cierta clase cuando, en realidad, en cada uno de los ítems solamente hay una única variable. El 6 % del alumnado, nombra las características de la variable sin utilizar expresiones incorrectas, aunque justifica sus respuestas sin el uso de términos específicos de estadística como cualitativo, cuantitativo, discreto o continuo (J3). Un 14 % del alumnado, a pesar de justificar su elección mediante argumentos estadísticos, utiliza expresiones incorrectas que pueden dar lugar a malentendidos como “variables independientes” o “cantidades exactas” (J4). Por último, se debe señalar la amplia proporción de alumnado que no muestra intención de ofrecer ningún tipo de argumento estadístico. Por un lado, el 19 % del alumnado justifica su elección con argumentos que no tienen base estadística (J5) como “más visual”, “fácil” o “comprensible”. Por otro lado, el 24 % ni siquiera responde.

Figura 13. Distribución de los niveles de justificación máximos del futuro profesorado



5. DISCUSIÓN Y CONCLUSIONES

En este artículo hemos analizado y evaluado el conocimiento que tiene el futuro docente de educación primaria sobre las medidas centrales y gráficos estadísticos. Más concretamente, hemos analizado si diferencia las categorías o valores de la variable estadística de las frecuencias absolutas en el cálculo de la moda, si hace una selección adecuada del tipo de gráfico según el tipo de variable, la corrección de los gráficos construidos y la justificación dada.

Observamos dificultades a la hora de distinguir entre los valores y las frecuencias absolutas, problema que también se ha detectado en varias investigaciones (Friel y Bright, 1996; Madrid et al., 2022 y Wu, 2004). A este resultado le podemos añadir que el alumnado también tiene dificultades para diferenciar las categorías y frecuencias absolutas de una variable cualitativa; así se refleja en el cálculo de la moda del ítem 1. Además, Groth y Bergner (2006) concluyeron que los futuros docentes de primaria relacionan el concepto de la moda únicamente con datos cuantitativos; precisamente, lo que manifiestan los resultados obtenidos en el ítem 1. Los futuros docentes de primaria tienen dificultades en comprender que la moda puede ser una respuesta no numérica, es decir, una categoría de la variable cualitativa. Las respuestas del ítem 1 también demuestran una confusión del alumnado entre la moda y media o mediana; justamente, lo opuesto a lo observado por Carvalho (2001), Freitas et al. (2018) y Leavy y O'loughlin (2006). Así, aunque son pocos los estudios que identifican dificultades en el cálculo de la moda, sí hay un porcentaje reseñable de estudiantes que dan respuestas erróneas.

Asimismo, hay un porcentaje elevado de estudiantes (28,7 %) que selecciona incorrectamente alguno de los gráficos. En este estudio, los errores más comunes han sido, por un lado, utilizar un gráfico de líneas para representar una variable cualitativa y, por otro lado, utilizar el histograma para una variable cualitativa o cuantitativa discreta sin agrupar, al igual que señalan Jacobbe y Horton (2010). Con respecto al primer error, debemos mencionar que los resultados aquí mostrados van en contra de lo que apuntan Alacaci et al. (2011), que perciben un dominio por parte del futuro profesorado para seleccionar el gráfico de líneas ante escenarios que describen tendencias. Un motivo puede ser el enfoque del estudio realizado, debido a que en la investigación de Alacaci et al. (2011), los gráficos están predeterminados y el alumnado sólo debe seleccionar uno, por lo que dicha elección se puede deber a un descarte. Con respecto al segundo error, tal como mencionan Arteaga et al. (2016), estamos de acuerdo en que su origen puede provenir de la aparente similitud con el diagrama de barras.

A la hora de construir un gráfico estadístico, los errores más destacados son no determinar los títulos o las etiquetas de los ejes y usar incorrectamente la escala de los mismos, coincidente con los ya identificados por Arteaga (2011) y Li y Shen (1992). En menor medida, también hemos observado errores de proporcionalidad entre los pesos relativos de los datos y su representación geométrica; error que también menciona Arteaga (2011). Otro error identificado es el intercambio en los ejes de los valores de la variable y las frecuencias absolutas, error ya detectado por Arteaga et al. (2011). Finalmente, Bruno y Espinel (2005) observan que es un error frecuente representar las barras no solapadas en los histogramas; en el presente estudio se detecta que el alumnado solapa las barras en los diagramas de barras.

Además, al igual que Alacaci et al. (2011) y Berciano et al. (2021), hemos comprobado cómo un porcentaje alto del alumnado no argumenta estadísticamente la elección de los gráficos usados, y solo en un porcentaje pequeño encontramos justificaciones correctas en el ámbito que les compete; todo ello plantea la duda sobre el interés que tiene el alumnado por la estadística y el uso correcto de la misma.

Finalmente, tras realizar un análisis pormenorizado de las dificultades que el futuro docente de primaria muestra en la realización de estas tareas básicas, estas podrían resumirse en: 1) Dificultad a la hora de identificar el tipo de variable: tendencia a clasificar cualquier variable estadística como cuantitativa discreta debido a que las frecuencias absolutas toman valores naturales. 2) Dificultad en el cálculo de la moda de variables cualitativas. 3) Dificultad en la selección del gráfico según el tipo de variable. 4) Confusión entre el diagrama de barras e histograma. 5) Ausencia de títulos o etiquetas, escalas o valores/categorías en los ejes. 6) Ausencia de valores/categorías o frecuencias absolutas/relativas en los diagramas de sectores. 7) Dificultades en la utilización del lenguaje estadístico: falta de precisión y confusión en la utilización de los términos *dato*, *variable estadística*, *valor* y *categoría*. 8) Falta de argumentación estadística. Así, para afrontar los errores identificados y poder mejorar la enseñanza-aprendizaje de la estadística con este colectivo sería interesante evaluar si el uso de contextos reales donde se deban interpretar gráficos de fuentes de información y comunicación influye en que el alumnado valore la importancia que tiene la presencia y corrección de los títulos o etiquetas y escalas. Otro punto sería investigar si hacer hincapié en la contextualización de los resultados obtenidos ayuda a no asociar un valor sin sentido a las medidas de centralización y dispersión, en general. Como futuro trabajo de investigación, sería interesante analizar la efectividad de dichas ideas para vencer las dificultades observadas.

6. AGRADECIMIENTOS

Artículo parcialmente financiado por el Grupo de Investigación GIU21/031 de la UPV/EHU.

REFERENCIAS

- Alacaci, C., Lewis, S.P., O'Brien, G.E., & Jiang, Z. (2011). Pre-Service Elementary Teachers' understandings of Graphs. *Eurasia Journal of Mathematics, Science and Technology Education*, 7(1), 3-14. <https://doi.org/10.12973/ejmste/75171>
- Alsina, A. (2012). La estadística y la probabilidad en Educación Infantil: conocimientos disciplinares, didácticos y experienciales. *Revista de Didácticas Específicas*, 7, 4-22. <https://revistas.uam.es/didacticasespecificas/article/view/7700/7976>
- Anasagasti, J. (2019). *Desarrollo de la competencia estadística del futuro docente de primaria a través del aprendizaje basado en proyectos*. Tesis Doctoral sin publicar, Universidad del País Vasco/Euskal Herriko Unibertsitatea. <https://addi.ehu.es/handle/10810/42760>
- Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores*. Tesis Doctoral sin publicar. Universidad de Granada. <https://www.ugr.es/~batanero/pages/ARTICULOS/arteaga.pdf>
- Arteaga, P., & Batanero, C. (2010). Evaluación de errores de futuros profesores en la construcción de gráficos estadísticos. En M. M. Moreno, A. Estrada, J. Carrillo & T. A. Sierra (Eds.), *Investigación en Educación Matemática XIV* (pp. 211-221). SEIEM. <https://www.seiem.es/docs/actas/14/Actas14SEIEM.pdf>
- Arteaga, P., Batanero, C. & Cañadas, G. (2011). Evaluación del conocimiento especializado de la estadística en futuros profesores en una tarea abierta. En M. Marín, G.

- Fernández, L. J. Blanco & M. Palarea (Eds.), *Investigación en Educación Matemática XV* (pp. 267-276). SEIEM.
- Arteaga, P., Batanero, C., Contreras, J. M., & Cañadas, G. (2016). Evaluación de errores en la construcción de gráficos estadísticos elementales por futuros profesores. *Revista Latinoamericana de Investigación en Matemática Educativa* 19(1), 15-40.
- Arteaga, P., Díaz-Levicoy, D., & Batanero, C. (2021). Primary school students' reading levels of line graphs. *Statistics Education Research Journal*, 20(2), Article 6. <https://doi.org/10.52041/serj.v20i2.339>
- Batanero, C., Arteaga, P., & Ruiz, B. (2009). Análisis de la complejidad semiótica de los gráficos producidos por futuros profesores de educación primaria en una tarea de comparación de dos variables estadísticas. *Enseñanza de las Ciencias*, 28(1), 141-154. <https://doi.org/10.5565/rev/ensciencias.3627>
- Batanero, C., Cobo, B., & Díaz, C. (2003). Assessing secondary school students' understanding of averages. *Proceedings of CERME III*. Bellaria, Italia.
- Batanero, C., & Godino, J.D. (2002). *Estocástica y su didáctica para maestros*. Universidad de Granada. https://www.ugr.es/~jgodino/edumat-maestros/manual/6_Estocastica.pdf
- Berciano, A., Anasagasti, J., & Zamalloa, T. (2021). Sentido estadístico en la formación de las y los estudiantes del Grado de Educación Infantil. Una aproximación desde un contexto de aprendizaje STEAM. *PNA*, 15(4), 289-309. <https://doi.org/10.30827/pna.v15i4.22510>
- Bruno, A., & Espinel, M. C. (2005). Recta numérica, escalas y gráficas estadísticas: un estudio con estudiantes para profesores. *Formación del Profesorado e Investigación en Educación matemáticas*, 7, 57-85. <http://fpiem.webs.ull.es/index.php/fpiem/article/view/18>
- Calot, G. (1974). *Curso de estadística descriptiva*. Paraninfo.
- Carvalho, C. (2001). *Interação entre pares. Contributos para a promoção do desenvolvimento lógico e do desempenho estatístico no 7º ano de escolaridade*. Tesis doctoral sin publicar, Universidad de Lisboa. <https://repositorio.ul.pt/handle/10451/42621>
- Cobo, B. (2003). *Significado de las medidas de posición central para los estudiantes de secundaria*. Tesis doctoral sin publicar. Universidad de Granada. <https://www.ugr.es/~batanero/pages/ARTICULOS/tesiscobo.pdf>
- Díaz-Levicoy, D. (2018). *Comprensión de gráficos estadísticos por alumnos chilenos de Educación Primaria*. Tesis doctoral sin publicar. Universidad de Granada. <https://digibug.ugr.es/bitstream/handle/10481/53598/29122880.pdf?sequence=4&isAllowed=y>
- Espinel, M.C. (2007). Construcción y razonamiento de gráficos estadísticos en la formación de profesores. En M. Camacho, P. Flores & M. Pilar Bolea (Eds.), *Investigación en Educación Matemática XI* (pp. 99-119). SEIEM. <https://www.seiem.es/docs/actas/11/Actas11SEIEM.pdf>
- Estrada, A. (2007). Actitudes hacia la estadística: Un estudio con profesores de Educación Primaria en formación y en ejercicio. En M. Camacho, P. Flores & M. Pilar Bolea (Eds.), *Investigación en Educación Matemática XI* (pp. 121-140). SEIEM. <https://www.seiem.es/docs/actas/11/Actas11SEIEM.pdf>
- Estrada, A., Batanero, C. & Fortuny, J.M. (2004). Un estudio sobre conocimientos de estadística elemental de profesores en formación. *Educación Matemática*, 16(1), 89-111. <https://doi.org/10.24844/EM1601.04>

- Fernandes, J. A. & Morais, P. C. (2011). Leitura e interpretação de gráficos estatísticos por alunos do 9º ano de escolaridade. *Educação Matemática Pesquisa*, 13(1), 95-115. <https://revistas.pucsp.br/index.php/emp/article/view/5282>
- Freitas, A., Figueiredo, T. S., Silva, N. & Miranda, M.C. (2018). Learning difficulties with median and quartiles concepts in the 8th grade students: formula versus plot comparative study. *Indagatio Didactica*, 10(2), 109-132. <https://doi.org/10.34624/id.v10i2.11313>
- Friel, S. N. & Bright, G. W. (1996). Building a theory of graphicacy: How do students read graphs? *Paper presented at the Annual Meeting of AERA*, New York.
- Groth, R. & Bergner, J. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median and mode. *Mathematical Thinking and Learning*, 8(1), 37-63. https://doi.org/10.1207/s15327833mtl0801_3
- Jacobbe, T. y Horton, R.M. (2010). Elementary school teachers' comprehension of data displays. *Statistics Education Research Journal*, 9(1), 27-45. [https://iase-web.org/documents/SERJ/SERJ9\(1\)_Jacobbe_Horton.pdf?1402525009](https://iase-web.org/documents/SERJ/SERJ9(1)_Jacobbe_Horton.pdf?1402525009)
- Leavy, A. M. & O'loughlin, N. (2006). Preservice teachers understanding of the mean: moving beyond the arithmetic average. *Journal of Mathematics Teachers Education*, 9, 53-90. <https://link.springer.com/content/pdf/10.1007/s10857-006-9003-y.pdf?pdf=button>
- Li, K. Y. & Shen, S. M. (1992). Students' weaknesses in statistical projects. *Teaching Statistics*, 14(1), 2-8.
- Madrid, A. E., Valenzuela-Ruiz, S. M., Batanero, C. & Garzón-Guerrero, J. A. (2022). Comprensión de la mediana por estudiantes universitarios. *AIEM-Avances de investigación en educación matemática*, 22, 1-21. <https://aiem.es/article/view/v22-madrid-valenzuela-batanero-et-al/3902-pdf-es>
- Mayén, S., Cobo, B., Batanero, C. & Balderas, P. (2007). Comprensión de las medidas de posición central en estudiantes mexicanos de bachillerato. *Revista Iberoamericana de Educación Matemática*, 3(9), 187-201. <https://union.fespm.es/index.php/UNION/issue/view/16/13>
- National Council of Teachers of Mathematics (NCTM). (2003). *Principios y estándares para la educación matemática*. Autor.
- Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria. <https://www.boe.es/buscar/pdf/2014/BOE-A-2014-2222-consolidado.pdf>
- Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. <https://www.boe.es/buscar/pdf/2022/BOE-A-2022-3296-consolidado.pdf>
- Sanoja de Ramirez, J. & Ortiz-Buitrago, J. (2013). Conocimiento de contenido estadístico de los maestros. En J. M. Contreras, G. R. Cañadas, M. M. Gea & P. Arteaga (Eds.), *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 157-164). Universidad de Granada. <https://www.ugr.es/~jmcontreras/pages/Investigacion/Actas%20jornadas.pdf>
- Wu, Y. (2004). Singapore secondary school students' understanding of statistical graphs. Trabajo presentado en el *10th International Congress on Mathematics Education*. Copenhagen, Dinamarca. <http://iase-web.org/documents/papers/icme10/Yingkang.pdf>

∞

Ane Izagirre

Universidad del País Vasco/Euskal Herriko Unibertsitatea (España)
ane.izagirre@ehu.eus | <https://orcid.org/0000-0001-8900-1576>

Jon Anasagasti

Universidad del País Vasco/Euskal Herriko Unibertsitatea (España)
jon.anasagasti@ehu.eus | <https://orcid.org/0000-0002-4732-7874>

Ainhoa Berciano

Universidad del País Vasco/Euskal Herriko Unibertsitatea (España)
ainhoa.berciano@ehu.eus | <https://orcid.org/0000-0001-7399-4745>

Recibido: 8 de julio de 2022

Aceptado: 3 de junio de 2023

Pre-service primary school teachers' statistical knowledge in data representation

Ane Izagirre @ , Jon Anasagasti @ , Ainhoa Berciano @ 

Universidad del País Vasco/Euskal Herriko Unibertsitatea (España)

In this article, we analyse if pre-service primary school teachers distinguish the categories or values of the statistical variable from the absolute frequencies when calculating the mode. Moreover, we investigate whether they select and construct the appropriate graph according to the type of variable and how they justify such choice. The work is carried out from an interpretative paradigm within mixed research, with a predominantly qualitative and exploratory nature, classifying and interpreting both the production of statistical graphs and the argumentation for their choice. The sample is conformed of 108 students in the third course of the Primary Education Undergraduate Degree of the University of the Basque Country. The instrument used for the collection of information is an ad-hoc questionnaire composed of 4 items. To analyse the information, two rubrics are presented. The first classifies the levels of correctness of the statistical graphs produced and has been compared with the work of Arteaga (2011). The second classifies the levels of justification for the selection of the graph and have been contrasted with the studies of Berciano et al. (2021) and Alacaci et al. (2011). In order to interpret the information a process of triangulation among researchers has been carried out. The results indicate that future teachers have difficulties in distinguishing the categories of a qualitative variable from its absolute frequencies. As regards the choice of graph, the most common errors are to make line graphs for qualitative variables and histograms for qualitative or discrete quantitative variables without grouping. The degree of correctness in making the graph and its justification is not sufficient, which shows the need to improve their statistical literacy. In conclusion, the following difficulties have been detected among pre-service primary school teachers: 1) Difficulty in identifying the type of variable: tendency to classify any statistical variable as discrete because absolute frequencies take natural values. 2) Difficulty in calculating the mode of qualitative variables. 3) Difficulty in the selection of the graph according to the type of variable. 4) Confusion between bar chart and histogram. 5) Absence of titles or labels, scales or values/categories on the axes. 6) Absence of values/categories or absolute/relative frequencies in the pie charts. 7) Difficulties in the use of statistical language: lack of precision and confusion in the use of the terms *data*, *statistical variable*, *value* and *category*. 8) Lack of statistical argumentation.