

Marco de desarrollo del razonamiento e instrumento de toma de datos sobre algunos conceptos básicos de muestreo

Felipe de Jesús Jiménez Rodríguez ¹
Ernesto Sánchez Sánchez ²

RESUMEN

En este artículo se presenta un marco de desarrollo del razonamiento sobre las nociones de población, muestra, muestreo repetido y distribución muestral, dirigido a estudiantes de bachillerato. El marco fue elaborado a partir de dos estudios empíricos que abarcan la distribución muestral, además de varios análisis de la literatura sobre muestras. Se presenta de manera esquemática en una tabla, donde se describen brevemente sus componentes. Asimismo, se incluye una descripción del instrumento utilizado para la recopilación de datos y su validación, llevada a cabo por tres expertos en educación estadística

PALABRAS CLAVE

Población, Muestra, Muestreo, Distribución muestral, Marco de desarrollo

¹ felipe.jimenez@cinvestav.mx

Departamento de Matemática Educativa, CINVESTAV, México
<https://orcid.org/0009-0000-0249-0347>

² esanchez0155@gmail.com

Departamento de Matemática Educativa, CINVESTAV, México
<https://orcid.org/0000-0002-8995-7962>

Jiménez Rodríguez, F. de J., & Sánchez Sánchez, E. (2024). Marco de desarrollo del razonamiento e instrumento de toma de datos sobre algunos conceptos básicos de muestreo. En M. Sánchez Aguilar, M. del S. García González, & A. Castañeda (Eds.), *Perspectivas actuales de la Educación Matemática* (pp. 135–145). Editorial SOMIDEM. <https://doi.org/10.24844/SOMIDEM/S3/2024/01-13>

Problema y pregunta de investigación

Los conceptos básicos de muestreo (en particular población, muestra y muestreo repetido) son necesarios para definir el concepto más avanzado de distribución muestral, el cual, a su vez, es crucial en la inferencia estadística. Como todos los conceptos, los de muestreo no son aprendidos por los estudiantes en su forma final ni de manera aislada, sino que los van captando de manera informal en diferentes etapas de aprendizaje, y están inmersos en procesos de razonamiento. En este trabajo sostenemos que el desarrollo del razonamiento inferencial se correlaciona con diferentes niveles de abstracción de los conceptos de muestreo. Hace falta, sin embargo, un marco que muestre de manera general cómo los estudiantes captan los conceptos de muestreo en diferentes etapas del desarrollo de su razonamiento inferencial. Nos preguntamos entonces: ¿Qué características tienen las versiones informales de los conceptos de muestreo para los estudiantes de nivel bachillerato y primer año de universidad?

Varios educadores estadísticos han propuesto marcos conceptuales o de desarrollo de conceptos fundamentales de la estadística (Jones, et al., 1997, Jones et al. 2000, Ben-Zvi, 2004, Reading & Reid, 2006, Reid & Reading, 2008, Noll & Shaughnessy, 2012). De manera similar, el objetivo del presente estudio es presentar un marco donde se formulan hipótesis acerca del desarrollo del razonamiento de estudiantes de bachillerato y primer año de universidad (de 16 a 18 años) en torno a los conceptos de población, muestra, muestreo y distribución muestral. Su elaboración se hizo teniendo en cuenta algunos resultados de estudios empíricos previos (García, 2017, , Silvestre et al., 2022) y el análisis de Ainley et al. (2015). El marco se representa esquemáticamente en la Tabla 1, donde se describen brevemente las categorías que lo forman con más detalle. En dicho marco se asume que las muestras son aleatorias. La evolución del razonamiento sobre los conceptos de aleatoriedad y representatividad no se incluyeron en el marco debido a que en los estudios referidos no se elaboraron instrumentos para recoger datos que dieran cuenta de las concepciones de los estudiantes acerca de ellos. Watson y Moritz (2000) ofrecen una descripción del desarrollo del concepto de muestra en niveles previos, en el que dan cuenta del papel de la representatividad.

Modelo del desarrollo del razonamiento

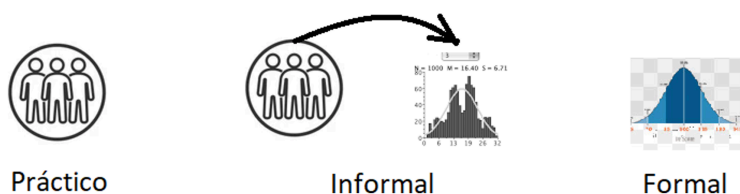
En general, llamamos razonamiento a los procesos mediante los cuales se obtiene y justifica una proposición (conclusión) a partir de datos y de otros conocimientos supuestos o dados (premisas), así, la función primordial del razonamiento es la justificación de proposiciones. Se formulan tres niveles que agrupan los diferentes razonamientos de los estudiantes en torno a los

conceptos elegidos; se definen de manera general de la siguiente manera (ver Figura 1):

- Entendemos un **razonamiento práctico** cuando las nociones involucradas en las premisas (frecuentemente implícitas) y en la conclusión se refieren a situaciones concretas.
- Un **razonamiento es informal** cuando incluyen representaciones simbólicas ligadas a las situaciones concretas, y son generadas con ayuda de conceptos y operaciones matemáticas.
- Un **razonamiento es formal** cuando se opera fundamentalmente con conceptos matemáticos independientemente de cualquier situación concreta.

Figura 1

Representación de los niveles de razonamiento práctico, informal y formal



Marco de desarrollo del razonamiento sobre conceptos de muestreo

La integración de los conceptos básicos de muestreo en conjunto con el modelo de razonamiento nos permite presentar de manera coherente y estructurada los elementos que conforman el Marco de Desarrollo del Razonamiento. Cada concepto dentro de este marco se manifiesta de forma distinta a medida que los estudiantes avanzan a través de los distintos niveles de razonamiento (ver Tabla 1). A continuación describimos los distintos niveles que pueden presentar los conceptos:

Población. En el nivel práctico, una población (1) son individuos u objetos concretos o materiales; en este nivel cae la definición de Moore (2000, p. 209): “Un grupo entero de individuos sobre los que queremos obtener información”. En el nivel informal se piensa una población (2) como representaciones numéricas o simbólicas (ocasionalmente organizadas en una distribución) obtenidas mediante una variable o atributo de un conjunto de individuos u objetos. En el nivel formal se concibe a una población (3) como una variable con una distribución teórica independiente de cualquier situación externa, pero que puede ser un modelo de una variable o atributo de un conjunto de individuos o cosas.

Muestra. Una muestra depende del concepto de población. En el nivel práctico, una muestra está formada de individuos u objetos, es “la parte de la

población que realmente examinamos con el objetivo de obtener información” (Moore, 2000, p. 209); esta caracterización resulta generalmente en una muestra sin reemplazo. En el nivel informal, una muestra es un subconjunto de los valores de una variable; esta concepción puede resultar tanto en una muestra sin reemplazo como en una con reemplazo. En el nivel abstracto, una muestra es el resultado de obtener valores de la distribución (sorteos de valores de la distribución, según Sóbol, 1976); en este caso, se obtiene una muestra con reemplazo.

Tabla 1

Marco de desarrollo del razonamiento sobre conceptos de muestreo

Conceptos	Nivel de razonamiento		
	Práctico	Informal	Abstracto o formal
Población	Población material: Conjunto de personas o cosas	Valores de una variable o atributo de individuos de una población material	Una variable y su distribución
Muestra	Subconjunto de la población material	Subconjunto de valores de una variable de la población	Una instancia de la n-éada de variables aleatorias X_1, X_2, \dots, X_n con $X_i \sim f(x)$, para toda i , y donde es una distribución (función de densidad de probabilidades)
Muestreo repetido	Obtener muestras reales sin reemplazo	Obtener muestras de valores de la variable	Sorteos o instancias del vector aleatorio: (X_1, X_2, \dots, X_n)
Distribución muestral (DM)	Representación de muestras reales	Representación de valores del estadístico de muestras reales	Distribución del estadístico: Distribución muestral Teórica

Muestreo repetido. En el nivel práctico, el muestreo repetido consiste en obtener muestras reales sin reemplazo (a nivel de individuos y de muestras) de una población material y finita, de esta manera, el agregado de las muestras resulta en una grande (Silvestre, et al. 2022). En el nivel informal, el muestreo repetido consiste en obtener muestras de valores de una variable o atributo de los individuos u objetos de muestras reales, pero el muestreo es con reemplazo a nivel de muestras. En el nivel abstracto, un muestreo es multiplicativo, y consiste en obtener muestras (matemáticas) de una distribución para determinar los patrones de variación muestral.

Distribución muestral (DM). En el nivel práctico, una DM es una representación de muchas muestras reales de una población; con esta concepción se cree que la DM se aproxima a la distribución de la población. En el nivel

informal, la DM es la distribución de los valores del estadístico de todas o muchas muestras de la población, a la que llamaremos Distribución Muestral Empírica (DME). En el nivel formal, una DM es una distribución teórica de probabilidad que modela el comportamiento de la variación muestral y es independiente de factores externos. Cuando el estadístico es la proporción de un atributo la DM, sólo depende de la hipótesis acerca de la proporción del atributo en una población y del tamaño de la muestra.

Es relevante destacar que Langrall et al. (2017) han enfatizado la necesidad de realizar estudios que prueben y validen los marcos de desarrollo, lo que permitiría diseñar la instrucción y la progresión curricular de manera más efectiva. En línea con esta perspectiva, este artículo incluye una descripción detallada del instrumento de recolección de datos utilizado para validar el marco propuesto.

Diseño del instrumento de recolección de datos para la validación del marco de desarrollo

El instrumento utilizado en esta investigación tiene como objetivo principal obtener información acerca del razonamiento de los estudiantes con relación al marco de la investigación. Con este propósito, se ha desarrollado un cuestionario que combina preguntas abiertas y de opción múltiple. Aunque por razones de espacio no es posible reproducir el cuestionario completo en este artículo, se encuentra disponible para su consulta en la siguiente liga: <https://rb.gy/9qefk>

Se propone utilizar para cuestionario el formato electrónico de los formularios de Google, con el objetivo de facilitar tanto la administración como el procesamiento de los datos recolectados. Esta elección presenta diversas ventajas, ya que permite la aplicación del cuestionario de manera amplia y generalizada, lo que facilita la obtención de una gran cantidad de datos. En contraste, el uso de protocolos de entrevistas o de observaciones en clase podría resultar limitado en términos de alcance y cantidad de información recopilada.

Para el análisis de las respuestas obtenidas se plantea un proceso de síntesis y comparación por cada ítem del cuestionario, con el fin de identificar patrones y categorías que puedan estar relacionadas con los conceptos del marco conceptual. Esta revisión permitirá la categorización de las respuestas en función de los diferentes niveles de desarrollo propuestos por el marco.

Descripción general del cuestionario

El cuestionario diseñado para esta investigación consta de 12 ítems distribuidos en cuatro grupos de tres ítems cada uno. Cada grupo está relacionado con una de las nociones fundamentales del marco de desarrollo del razonamiento: población, muestra, muestreo repetido y distribución muestral.

En cada grupo de ítems se presentan dos tipos de preguntas. El primer tipo de pregunta es de respuesta abierta y busca indagar el conocimiento que los participantes tienen sobre el concepto en cuestión, permitiéndoles expresar sus ideas sin restricciones. Luego se presentan dos preguntas de opción múltiple (con cuatro o cinco opciones de respuesta), en las que se solicita a los participantes que seleccionen la opción que consideren más apropiada y que justifiquen su elección o, si lo consideran pertinente, propongan una alternativa adicional.

Tipos de preguntas utilizadas en el cuestionario

El grupo I de ítems se diseñó con el propósito de indagar en qué nivel de razonamiento se encuentra la concepción de población de los participantes. En particular, observemos la estructura del ítem 2 (ver tabla 2). El inciso a del ítem presenta una respuesta que refleja un nivel de razonamiento práctico, ya que implica relacionar a los estudiantes con la población, asumiendo una perspectiva material del concepto. Por otro lado, los incisos b y c ofrecen respuestas que se sitúan en un nivel de razonamiento informal, ya que incluyen una representación simbólica, en este caso un conjunto, de la posible población, pero aún no consideran la probabilidad asociada a cada elemento. Finalmente, el inciso d presenta una expresión más formal de la población asociada dentro del marco conceptual establecido, ya que contempla los posibles resultados del experimento y considera las probabilidades al mencionar la distribución. Es importante precisar que los incisos a y b no son aceptables, mientras que los incisos c y d son posibles normativamente.

Tabla 2

Ítem 2 del cuestionario para la toma de datos

-
1. **En una escuela se realiza un concurso de tiros libres a la canasta de basquetbol. Cada estudiante lanza 10 tiros y se registra el número de tiros encestandos. ¿Cuál de las siguientes opciones elegirías como la población del experimento anterior?**
 - a) Los estudiantes de dicha escuela.
 - b) El conjunto $I = \{0, 1\}$, donde 1 = "Encestar", 0 = "No encestar".
 - c) El conjunto de todas las secuencias o arreglos de "0" y "1" de tamaño 10. Ejemplo de una de esas secuencias: (1,0,1,0,1,0,0,0,1,0,1).
 - d) Los valores de la variable: "El número de tiros anotados en 10 lanzamientos". Es decir, $A = \{0, 1, 2, \dots, 10\}$ y su distribución.
 - e) Otra. Explica tu respuesta.

Por favor, explica a detalle por qué elegiste la opción de la pregunta anterior.

(Respuesta abierta)

Conviene señalar que el inciso d se encuentra asociado con una población cuya distribución es binomial. Esta distribución se obtiene a partir de una variable aleatoria que registra 10 eventos independientes. No obstante, esta característica no se explica detalladamente en aras de evitar una formulación demasiado abstracta.

El grupo II de ítems se relaciona con la noción de muestra y presenta una estructura similar a la del grupo I. En este caso se formula una pregunta abierta para saber el conocimiento que el estudiante atribuye a dicha noción, seguida de dos preguntas de opción múltiple que permiten ubicar su respuesta dentro de alguno de los niveles del marco. Por ejemplo, en el ítem 6 del grupo II, (ver tabla 3) el inciso a se representa una muestra en el nivel práctico, mientras que el inciso b refleja un nivel de razonamiento informal debido a que la muestra presenta un componente simbólico, en este caso, una representación en forma de arreglo o secuencia. Por su parte, el inciso c presenta una muestra que surge de asociar el experimento de la obtención de las fichas de la urna con una distribución Bernoulli, aunque esta relación no se explica de manera explícita. En este caso, las opciones a, b y c son aceptables desde un punto de vista normativo.

Cabe destacar que la finalidad de este conjunto de ítems sobre muestras es identificar si los estudiantes las conciben como subconjuntos de la población o si, por el contrario, las conciben de manera más abstracta como sorteos de valores de la variable asociada a la población del problema. Es importante mencionar que el contexto de los ítems se mantiene constante en todos los grupos considerados.

El grupo III de ítems se centra en la noción del muestreo repetido, otorgándole especial atención a la elección entre muestreo repetido sin reemplazo o muestreo repetido con reemplazo. Como se mencionó en la descripción del marco conceptual, la selección de un tipo de muestreo repetido conlleva implicaciones en el nivel de abstracción del muestreo que considera el estudiante.

Por otro lado, el grupo IV de ítems se enfoca en la obtención de información acerca del concepto de distribución muestral. Para lograr este objetivo, se plantearon dos ítems de preguntas que abordan aspectos importantes relacionados con dicho concepto. El ítem 11 indaga acerca de la representación gráfica que los participantes consideran adecuada para un muestreo repetido de una gran cantidad de muestras extraídas de una distribución inicial. En las opciones se presentan, en el nivel práctico, distribuciones que tiene una forma parecida a la de la población, y en los otros niveles, distribuciones que tiene forma normal. Por su parte, el ítem 12 requiere que se identifique la inusualidad de una muestra con base en una distribución inicial.

Evaluación de expertos sobre el instrumento de toma de datos

En el proceso de validación del instrumento de recolección de datos, se contó con la participación de tres expertos en educación estadística, quienes desempeñaron el papel de evaluadores. Se les solicitó su opinión sobre la formulación, pertinencia y relevancia de los 12 ítems propuestos en el cuestionario dirigido a los alumnos, con relación al marco de desarrollo.

Para recabar sus opiniones, se diseñó un sondeo específico que se les envió por correo electrónico. Las respuestas de los expertos fueron recopiladas para un posterior análisis comparativo (ver tabla 4).

Tabla 3

Ítem 6 del cuestionario para la toma de datos

1. **Se quiere obtener una muestra de tamaño 10 de una urna con 500 fichas blancas y 800 fichas negras ¿Cuál de los siguientes incisos elegirías como muestra?**
 - a) Diez fichas sacadas de la urna.
 - b) Una secuencia o arreglo de tamaño 10 compuestos por ceros (0) y unos (1), donde "0" significa "ficha negra" y "1" significa "ficha blanca". Ejemplo de una secuencia: (0, 1, 0, 1, 1, 0, 1, 1, 0, 1).
 - c) Diez valores aleatorios de la variable "El resultado extraer una única ficha de una urna que solo tiene fichas negras y blancas". Es decir, $B = \{0, 1\}$, donde la probabilidad de sacar una ficha blanca (1) es de 500/1300, mientras que la probabilidad de sacar una ficha negra (0) es de 800/1300.
 - d) Otra. Explica tu respuesta en la siguiente pregunta.

Por favor explica a detalle por qué elegiste la opción de la pregunta anterior.
(Respuesta abierta)

Tabla 4

Comparativa de las valoraciones de los evaluadores

Conceptos del Marco de Desarrollo Del Razonamiento												
	Población			Muestra		Muestreo Repetido			Distribución Muestral			
<i>Ítem</i>	1	2	3	4	5	6	7	8	9	10	11	12
Evaluador 1	✓	?	✓	✓	✓	?	✓	?	✓	✓	✓	X
Evaluador 2	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	?
Evaluador 3	✓	✓	?	✓	✓	?	✓	✓	✓	✓	✓	?

Valoración general: Los evaluadores consideraron que el cuestionario es aplicable con correcciones menores

Tipos de valoraciones: ✓: No propone modificaciones; ?: Propone alguna modificación; X: No aprueba el ítem.

Conceptos del Marco de Desarrollo Del Razonamiento												
	Población			Muestra		Muestreo Repetido			Distribución Muestral			
<i>Ítem</i>	1	2	3	4	5	6	7	8	9	10	11	12
Evaluador 1	✓	?	✓	✓	✓	?	✓	?	✓	✓	✓	X
Evaluador 2	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	?
Evaluador 3	✓	✓	?	✓	✓	?	✓	✓	✓	✓	✓	?

Valoración general: Los evaluadores consideraron que el cuestionario es aplicable con correcciones menores

Tipos de valoraciones: ✓: No propone modificaciones; ?: Propone alguna modificación; X: No aprueba el ítem.

En términos generales, los tres evaluadores coincidieron en que el cuestionario es aplicable con modificaciones menores. Las respuestas a cada valoración de los ítems se clasificaron en tres categorías: "no propone modificaciones", "propone alguna modificación" y "no aprueba el ítem". La tabla 4 presenta el cuadro comparativo resultante de esta clasificación.

De los 12 ítems evaluados, en 27 de las 36 evaluaciones se consideró que no era necesario realizar modificaciones. En 8 respuestas, los evaluadores sugirieron algún cambio en los ítems. Únicamente en el caso del ítem 12, el evaluador 1 argumentó que este carecía de pertinencia con relación a los conceptos del marco y recomendó su reemplazo por otro tipo de ítem.

Reflexión final

En el marco presentado se propone la idea de que un razonamiento informal sobre los términos población, muestra, y muestreo repetido consiste en una instancia intermedia entre concepciones de ellos basados en razonamientos con elementos concretos, y argumentaciones matemáticas abstractas. Así, lo informal se presenta como una etapa de transición entre el mundo sensible y el universo matemático. De lo anterior se deriva que el razonamiento sobre una distribución muestral también tiene una versión concreta, una informal y una matemática. Esta observación aún no se ha tenido en cuenta en la literatura didáctica sobre el tema; su reconocimiento y aplicación en la educación estadística puede ayudar a entender algunas de las dificultades de los estudiantes con el concepto de distribución muestral, y así abrir nuevas vías para su investigación. También puede ser útil para su aplicación práctica en el aula.

El diseño del instrumento de recolección de datos se compone de un cuestionario que tiene como objetivo indagar acerca de las cuatro nociones del marco sobre muestreo que hemos destacado. A través de la aplicación de este cuestionario, se pretende clasificar las respuestas de los estudiantes con base en los niveles de razonamiento establecidos en el modelo de razonamiento. Para asegurar la calidad y validez del instrumento, se llevó a cabo un sondeo en el que participaron tres expertos en educación estadística, quienes evaluaron los ítems del cuestionario. Los evaluadores coincidieron en que el cuestionario dirigido a los alumnos es aplicable y pertinente, requiriendo solo modificaciones menores.

Enseguida describiremos algunas de las limitaciones de este estudio. En primer lugar, es importante destacar que tanto el marco conceptual como el cuestionario para alumnos pueden expandirse para abordar las ideas de aleatoriedad, variabilidad, distribución y probabilidad, ya que estos conceptos juegan un papel importante en la construcción del concepto de distribución muestral. La segunda limitación es que el cuestionario utilizado en esta investigación, si bien proporciona información acerca de las concepciones

generales y el nivel de abstracción de los estudiantes con relación al marco conceptual, no permite obtener datos detallados sobre el conocimiento técnico y operativo de los conceptos específicos. En tercer lugar, considerando que cada estudiante puede poseer una perspectiva única y personal con relación a los conceptos estudiados, es probable que algunas de sus respuestas no se ajusten perfectamente a las categorías predefinidas o esperadas; esto se debe a la diversidad en los conocimientos individuales y a la forma en que cada estudiante ha internalizado dichos conceptos. Por lo tanto, los niveles que se proponen describen ciertas tendencias generales, y en cada una de ellas caben variaciones particulares.

Finalmente, la investigación se encuentra en curso con el propósito de avanzar en estos aspectos y lograr un mayor nivel de profundidad y precisión en la evaluación del conocimiento de los estudiantes.

REFERENCIAS

- Ainley, J., Gould, R., & Pratt, D. (2015). Learning to reason from samples: Commentary from the perspectives of task design and the emergence of 'Big data'. *Educational Studies in Mathematics* 88(3), 405–412. <https://doi.org/10.1007/s10649-015-9592-4>
- Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42–63. <https://doi.org/10.52041/serj.v3i2.547>
- García, V. (2017) *Diseño de una Trayectoria Hipotética de Aprendizaje para la Introducción y Desarrollo del Razonamiento sobre el Contraste de Hipótesis en el Nivel Medio Superior*. [Tesis doctoral no publicada, Cinvestav].
- Jones, G. A., Langrall, C. W., Thornton C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics* 32(2), 101–125. <https://doi.org/10.1023/A:1002981520728>
- Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning*, 2(4), 269–307. https://doi.org/10.1207/S15327833MTL0204_3
- Langrall, C.H., Makar, K., Nilsson, P., & Shaughnessy, J. M. (2017). Teaching and Learning probability and statistics: An integrated perspective. En J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 490–52. National Council of Teachers of Mathematics.
- Moore, D. S. (2000). *Estadística Aplicada Básica*. Antoni Bosch Editor.
- Noll, J., & Shaughnessy, M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. *Journal for Research in Mathematics Education*, 43(5), 509–556. <https://doi.org/10.5951/jresmetheduc.43.5.0509>

- Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal*, 5(2), 46–68. <https://doi.org/10.52041/serj.v5i2.500>
- Reid, J., & Reading, C. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal*, 7(1), 40–59. <https://doi.org/10.52041/serj.v7i1.479>
- Silvestre C., Sánchez, E., & Inzunza, S. (2022). El razonamiento de estudiantes de bachillerato sobre el muestreo repetido y la distribución muestral empírica. *Educación Matemática*, 34(1), 100–130. <https://doi.org/10.24844/em3401.04>
- Watson, J. M., & Moritz, J. B. (2000). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, 19(1) 109–136. [https://doi.org/10.1016/S0732-3123\(00\)00039-0](https://doi.org/10.1016/S0732-3123(00)00039-0)