ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

Niveles de Razonamiento Inferencial para el Estadístico t-Student

Inferential Reasoning Levels for t-Student Statistical

Jesús Guadalupe **Lugo-Armenta**[*]

ORCID iD 0000-0001-6679-5115

Luis R. **Pino-Fan**[**]

ORCID iD 0000-0003-4060-7408

Resumen

En este artículo presentamos una propuesta de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial para el estadístico t-Student, a partir de criterios epistémicos identificados con un estudio de tipo histórico-epistemológico sobre este estadístico y de la investigación desarrollada sobre razonamiento inferencial. Para ello, utilizamos algunas nociones teórico-metodológicas introducidas por el Enfoque Onto-Semiótico del conocimiento y la instrucción matemáticos (EOS), las cuales permitieron tanto identificar y caracterizar diversos significados conferidos al estadístico t-Student, a lo largo de su evolución y desarrollo, como presentar una perspectiva integral de lo que se considera razonamiento inferencial. Los atributos matemáticos de los diversos significados del estadístico t-Student se encuentran fuertemente vinculados a los indicadores de los distintos niveles de razonamiento aquí expuestos. Además, cada nivel se encuentra asociado a un razonamiento inferencial informal, pre-formal o formal. La propuesta de niveles de Razonamiento Inferencial para el estadístico t-Student y sus indicadores, se prevén útiles para el diseño de actividades que promuevan, gradualmente, un razonamiento inferencial formal sobre la base del razonamiento inferencial informal, sobre este estadístico.

Palabras clave: Educación Estadística. Razonamiento inferencial. Inferencia estadística. t-Student.

Abstract

In this article, we present a proposal of progressive inferential reasoning levels, from the informal to the formal, for the t-Student statistic, based on epistemic criteria identified with a historical-epistemological study on this statistic and from the investigation developed on inferential reasoning. To do this, we used some theoretical-methodological notions introduced by the Onto-Semiotic Approach to mathematical knowledge and instruction (OSA), which allowed both to characterize various meanings conferred on the t-Student statistic throughout its evolution and development, and to present a comprehensive perspective of what is considered inferential reasoning. The mathematical attributes of the t-Student statistics' meanings are strongly linked to the indicators of the different levels of inferential reasoning presented here. Furthermore, each level is associated with informal, pre-formal, or formal inferential reasoning. The proposed levels of Inferential Reasoning for the t-Student statistic and its indicators are expected to be useful for the design of activities that gradually promote formal inferential reasoning based on informal inferential reasoning on this statistic.

Keywords: Statistical Education. Inferential reasoning. Statistical inference. t-Student.

[*] Doctor en Educación Matemática por la Universidad de Los Lagos (ULAGOS). Académico del Departamento de Ciencias Exactas de la Universidad de Los Lagos (ULAGOS), Osorno, Chile. E-mail: jesus.lugo@ulagos.cl.
[**] Doctor en Didáctica de la Matemática por la Universidad de Granada (UGR). Académico del Departamento de Ciencias Exactas de la Universidad de Los Lagos (ULAGOS), Osorno, Chile. E-mail: luis.pino@ulagos.cl.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

1 Antecedentes

En Latinoamérica, por muchos años se pensó que la inferencia estadística se encontraba únicamente en los currículos universitarios, pero recientemente ha habido cambios respecto de esta visión. Diversas investigaciones (e.g., CALLINGHAM; WATSON, 2017; PFANNKUCH, 2018), han evidenciado que existen rasgos de la inferencia estadística en años previos a la educación universitaria, con lo cual los currículos de matemáticas, por ejemplo el chileno, contemplan promover que los estudiantes tomen decisiones y las fundamenten a partir del análisis crítico de la información que se les proporciona y modelos de probabilidad (e.g., MINEDUC, 2019).

Este cambio – motivado principalmente por la importancia que tiene la inferencia estadística para desarrollar un tipo de razonamiento en las personas que les permita entender el mundo en el que viven, tomando decisiones a partir de análisis estadísticos y la interpretación critica de la información, tanto en la vida cotidiana como laboral – ha generado un creciente interés por parte de la comunidad de investigación, para intentar comprender qué es el razonamiento inferencial (RI) y buscar formas de aproximarnos a su desarrollo.

En este sentido, han surgido diversas propuestas para caracterizar y desarrollar el razonamiento inferencial, desde propuestas que apuntan a un razonamiento inferencial informal – RII – (e.g., DOERR; DELMAS; MAKAR, 2017; MAKAR; RUBIN, 2009; ZIEFFLER *et al.*, 2008), hasta aquellas que proponen el desarrollo de un razonamiento inferencial formal – RIF – a partir de una progresión desde el RII (e.g., ARNOLD; WILD, 2015; JACOB; DOERR, 2014; PFANNKUCH; MAKAR; RUBIN, 2018). Sin embargo, aún no hay un consenso sobre cómo podemos construir un RIF sobre las bases que se han desarrollado de un RII.

En este artículo, se toma como supuesto que, para promover progresivamente un RIF, es necesario conocer cómo las nociones estadísticas emergen de las prácticas matemáticas que permitieron resolver distintos tipos de problemas, esto nos permitiría identificar los diversos significados de la misma noción. En este estudio, tomamos como ejemplo el estadístico t-Student, por su importancia en la aplicación de la Estadística Inferencial y en la Educación Estadística, e identificamos los diversos significados que ha adquirido a lo largo de su evolución y desarrollo histórico. Los significados permitieron obtener criterios epistémicos para la construcción de niveles de razonamiento inferencial para este estadístico.

Una de las bondades características de las pruebas con el estadístico t-Student es que pueden ser utilizadas para realizar inferencias cuando se tienen muestras pequeñas. En los estudios de áreas como la industria automotriz, farmacéutica y de la agronomía, así como

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

estudios sociales, es muy común contar con ensayos o participantes limitados, en cuyos casos son esenciales las pruebas con el estadístico t-Student. Sin embargo, realizar inferencias basados en dichas pruebas requiere una comprensión profunda tanto del estadístico t-Student, como de las nociones que se encuentran relacionadas con él, por ejemplo, el nivel de significancia, las hipótesis estadísticas, el valor-p y las distribuciones muestrales; mismas nociones con las que se han identificado dificultades al realizar inferencias (e.g. GARFIELD; BEN-ZVI, 2008; HARRADINE; BATANERO; ROSSMAN, 2011; SOTOS *et al.*, 2007).

Así, el objetivo de este artículo es presentar una propuesta teórica de niveles progresivos de razonamiento inferencial, de lo informal a lo formal, sobre el estadístico t-Student, tomando como base tanto la riqueza matemática que se recupera del estudio histórico-epistemológico sobre este estadístico, como los aportes de la literatura científica de Educación Estadística sobre el razonamiento inferencial. Los niveles de RI propuestos presentan *indicadores* graduables en procesos de generalidad y formalización, que permiten diseñar actividades que promuevan el razonamiento inferencial de forma progresiva, así como, también, realizar análisis detallados de las prácticas matemáticas (del currículo, los libros de texto, de estudiantes etc.) para determinar niveles de RI promovidos.

2 Marco Teórico-Metodológico

Para el desarrollo de esta investigación utilizamos algunas de las nociones teórico-metodológicas del Enfoque Ontosemiótico (EOS) del conocimiento y la instrucción matemáticos (GODINO; BATANERO; FONT, 2007, 2019). 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 cobra un rol fundamental en el EOS, entendiéndolas 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; BATANERO, 1994, p. 334).

Las prácticas matemáticas pueden ser personales (prácticas personales) o compartidas por un grupo en una institución (prácticas institucionales). Entonces, en estas prácticas operativas y discursivas que realizan las personas al resolver un cierto tipo de problemas, intervienen y emergen objetos matemáticos (ostensivos y no ostensivos) que son representados en forma textual, oral, gráfica o incluso gestual. Además, de los sistemas de prácticas matemáticas, operativas y discursivas emergen, al menos, seis nuevos objetos, denominados objetos matemáticos primarios: elementos lingüísticos (representaciones),

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

situaciones/problemas, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos (GODINO; BATANERO; FONT, 2007, 2019). Estos objetos matemáticos primarios están relacionados entre sí configurando la actividad matemática.

En este sentido, el *significado* de un objeto matemático es entendido en el EOS como el sistema de prácticas operativas y discursivas que una persona o una institución desarrolla para resolver cierto tipo de situaciones/problemas en los que interviene dicho objeto matemático (GODINO; BATANERO, 1994). Así, la noción de significado se concibe desde una perspectiva pragmático-antropológica que considera la relatividad del contexto en el que se utilizan los objetos matemáticos. Entonces, al estudiar ¿cuál es el significado del estadístico t-Student?, analizamos tanto las prácticas de las cuales emergió esta noción, como aquellas que permitieron su evolución y desarrollo, identificando los contextos que atiende.

Además, los objetos matemáticos primarios pueden ser analizados desde una perspectiva proceso-producto en el EOS, algunos procesos que permiten entender la naturaleza compleja y progresiva de los objetos matemáticos son: generalización, particularización (ejemplificación), idealización (esquematización), materialización, representación, significación, reificación, descomposición, modelización (FONT; RUBIO, 2017; PINO-FAN; GODINO; FONT, 2018).

Con base en lo anterior, podríamos decir que en el EOS el *razonamiento* se asume como un *macro proceso social y epistémico*, que involucra poner en juego tanto los objetos matemáticos primarios como los procesos señalados anteriormente, para la solución de una situación-problema (e.g., MOLINA, 2019; AKÉ, 2013). Es decir, un sujeto evidencia su nivel de razonamiento sobre la t-Student en la medida que en sus prácticas (para resolver distintos tipos de situaciones/problemas), emerjan de manera gradual, sistemática y progresiva, objetos matemáticos primarios y procesos vinculados a los significados de esta noción.

3 Significados del Estadístico t-Student

Mediante el estudio histórico-epistemológico que se realizó para el estadístico t-Student, se identificaron tres grandes problemáticas que resultaron clave para el surgimiento, desarrollo y generalización de este estadístico. Estas problemáticas, que constituyen tres grandes significados para el estadístico, son: (1) prueba de la media de una muestra, (2) distribución t-Student, y (3) prueba de la diferencia de medias para dos muestras. En las prácticas matemáticas que se llevaron a cabo para resolver estas problemáticas, fue posible identificar doce significados (parciales) donde se movilizaron distintas situaciones-problema, elementos

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

lingüísticos, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos. Estos doce significados parciales (SP) constituyen el significado holístico del estadístico t-Student (Figura 1).

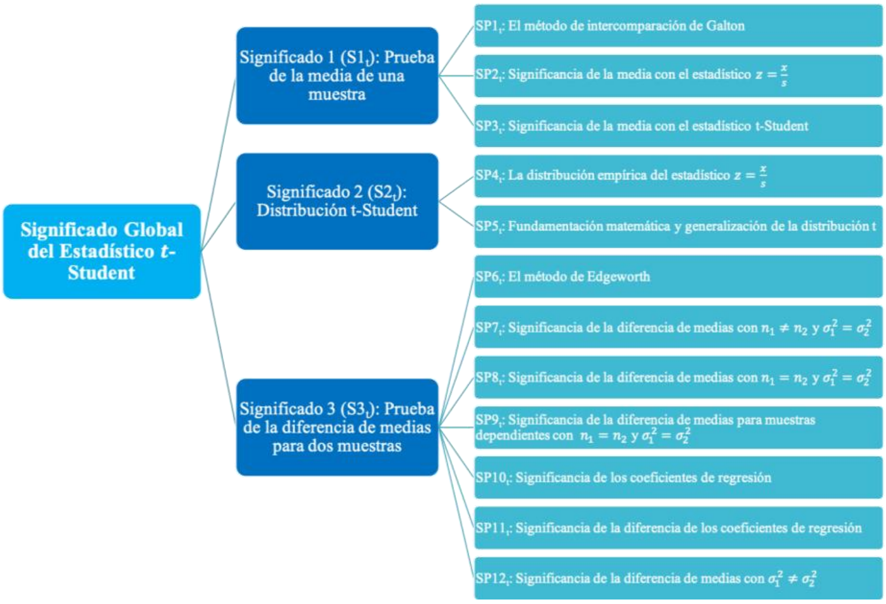

Figura 1 – Significados del estadístico t-Student
Fuente: elaborado por el autor

La prueba de la media de una muestra ($S1_t$) nos permite determinar si la media de una población, que se distribuye de acuerdo a la normal, es igual a un valor μ_0. La interrogante ¿cómo se distribuye el estadístico t-Student?, constituyó el problema principal que detonó el $S2_t$. Mientras que en la prueba de la diferencia de medias para dos muestras ($S3_t$) se encuentra la generalización de algunos elementos del $S1_t$, con esta prueba podemos determinar si las medias de dos poblaciones distribuidas de forma normal son iguales, en otras palabras, si no existe diferencia significativa entre las medias.

El significado parcial uno y seis corresponden a una versión intuitiva de estas pruebas, por ejemplo, en el $SP1_t$ se puede analizar la variación interna de un conjunto de datos bajo un método de comparación donde se hace uso de percentiles y cuartiles, principalmente, estas medidas se determinan al ordenar y graficar los datos; a este método Galton (1875) lo denominó intercomparación. Gosset (STUDENT, 1908) dio un uso más matematizado a la intercomparación dando paso al $SP2_t$. Mientras que en el $SP6_t$ se analiza la fluctuación como una medida de la variación de una serie de datos con respecto a otra serie (EDGEWORTH, 1885), este método sigue la idea de Galton, sobre que las comparaciones se pueden realizar únicamente en términos de la variación interna del conjunto de datos y no tomando como referencia algún criterio externo. Los $SP3_t$, $SP5_t$, $SP7_t$, $SP8_t$, $SP9_t$, $SP10_t$ y $SP11_t$ dan cuenta de la prueba, generalización y extensión del estadístico t-Student que realizó Fisher.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

4 Niveles de Razonamiento Inferencial para el Estadístico t-Student

A partir de los aportes de la literatura científica en didáctica de la estadística y de los objetos matemáticos primarios (elementos lingüísticos, conceptos/definiciones, propiedades, procedimientos y argumentos) y procesos identificados en cada uno de los significados parciales resultantes del estudio histórico-epistemológico, fue posible caracterizar cuatro niveles de razonamiento inferencial para la t-Student. Iremos haciendo énfasis a los objetos matemáticos primarios y procesos asociados a cada uno de los significados del estadístico, conforme presentemos los niveles y sus descriptores.

4.1 Nivel 1

Los indicadores de este nivel corresponden a un razonamiento inferencial informal, éstos van más allá de simples interpretaciones de los gráficos o cálculos de medidas como la media, están enfocados en razonar inferencialmente y realizar conjeturas a partir de los datos. Los argumentos que apoyan las conjeturas están basados en la variación interna de los datos; inicialmente, dicha variación es observable en las gráficas y, posteriormente, en el análisis que se realiza de una o dos muestras. Los objetos matemáticos que intervienen para analizar los datos, así como para establecer y argumentar las conjeturas pertenecen a Estadística Descriptiva y Probabilidad.

Sub-nivel 1.1. Visualización

Se espera que el estudiante sea capaz de identificar la variación interna de un conjunto de datos a través de observar e interpretar la forma y los cuartiles de la gráfica (e.g., Figura 2), y razone sobre cómo podría comportarse la población de donde se extrajeron estos datos.

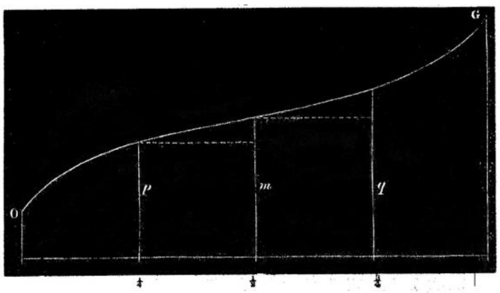

Figura 2 – Ojiva resultante del método de intercomparación
Fuente: GALTON (1875, p. 36)

Los elementos matemáticos que se movilizan en este nivel son del $SP1_t$, por ejemplo: cuartiles, valor medio y la simetría de la serie.

Sub-nivel 1.2. Trabajar con datos de una muestra

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

El estudiante puede utilizar el método de intercomparación, donde identifica valores representantes (e.g., cuartiles, octiles, deciles y percentiles); realiza la gráfica correspondiente y compara e interpreta la dispersión de los intervalos, los intervalos son formados por valores representantes, como los cuartiles (e.g., Figura 3).

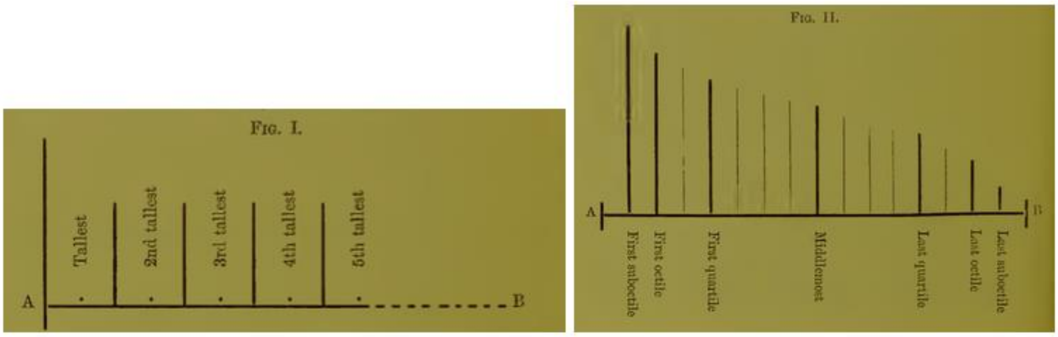

Figura 3 – Gráficos del método de intercomparación con diversos representantes
Fuente: GALTON (1880, p. 8)

Los conceptos/definiciones que se movilizan son error probable como una medida de variabilidad de la serie observada, ojiva, cuartiles, valor medio y percentiles; mientras que ejemplo de propiedades/proposiciones son m en $\frac{1}{2}$ o $0°$, p en $\frac{1}{4}$ o $-25°$, q en $\frac{3}{4}$ o $25°$, error probable como $q - m$ y $m - p$, y la propiedad sobre la simetría de la serie como $q - m = m - p$; todos ellos correspondientes al $SP1_t$, el cual surge a partir de los trabajos de intercomparación de Galton (1875).

Es importante que, posteriormente, el estudiante establezca una conjetura de la población fundamentada en la interpretación de los gráficos y en la dispersión de los intervalos que ha calculado.

Sub-nivel 1.3. Trabajar con datos de dos muestras

El estudiante puede establecer una conjetura argumentada en la variación interna de las dos muestras. Para lograrlo:

- Debe ser capaz de analizar, bajo el método de intercomparación, cada una de las muestras y comparar los valores representantes (e.g., cuartiles) de cada una de las muestras, puede realizar gráficos (e.g., diagrama de caja y bigote) y conjetura las semejanzas y/o diferencias entre las muestras. Esto se realiza a manera de conexión con el nivel 1.2 y, posteriormente, dar paso al siguiente indicador.
- Puede analizar la variación interna de los datos a través de las gráficas que elabora (e.g., diagramas de caja y bigote, y ojiva) y de la fluctuación $\left(2\frac{\sum e^2}{n}\right)$, donde n es la suma de las diferencias $(y_i - x_i)$ y $e = (y_i - x_i) - (\bar{y} - \bar{x})$.

En este sub-nivel intervienen objetos matemáticos primarios del $SP6_t$ como los

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

conceptos/definiciones de población, media y fluctuación entendida como una medida de la variación de una serie de datos con respecto a otra serie; mientras que algunas propiedades/proposiciones que intervienen son $\bar{x} = \frac{1}{n}\sum x_i$ [1] y fluctuación como $2\frac{Se^2}{n}$, donde n es la sumatoria de las diferencias. La noción de fluctuación se rescató del método de análisis para tablas estadísticas que realizó Edgeworth (1885), este análisis se basa solamente en la variación interna de las muestras.

En el nivel 1 podemos encontrar características del marco de trabajo de Zieffler *et al.* (2008), como es el caso de los tres componentes del RII (hacer juicios o predicciones, usar o integrar el conocimiento previo y articular evidencia basada en argumentos), y también la similitud con las tareas 1 y 2 (estimar y esbozar un gráfico de la población y comparar dos muestras de datos). En diversas investigaciones (e.g., BAKKER; BEN-ZVI; MAKAR, 2017; BEN-ZVI; ARIDOR-BERGER, 2016; MAKAR; BEN-ZVI, 2011) se ha evidenciado la importancia de trabajar con situaciones en contextos cercanos a los estudiantes, y en especial al promover un RII, ya que este tipo de razonamiento no contempla únicamente el conocimiento estadístico de los estudiantes sino también su conocimiento informal.

Al abordar los indicadores de este nivel es importante que el estudiante también razone sobre ideas como variabilidad y distribución, error, muestreo e inferencia y predicción; estas ideas se han promovido dentro de la inferencia informal en diversas investigaciones (e.g., DE VETTEN *et al.*, 2018; GARFIELD *et al.*, 2015; NOLL; HANCOCK, 2015; PFANNKUCH, ARNOLD; WILD, 2015).

4.2 Nivel 2

En este nivel se encuentran indicadores que corresponden a un nivel de tipo pre-formal, ya que podemos identificar rasgos del RII tales como la forma en que se utiliza la probabilidad, como una medida de certidumbre asociada a un evento sin compararla con un nivel de significancia o un límite preestablecido, y cómo se observa la hipótesis nula enmarcada dentro del problema (se puede considerar una primera aproximación a la hipótesis, ya que esta se encuentra implícita). Esta forma de utilizar la probabilidad se encuentra en los inicios (histórico) de las pruebas t-Student.

Sub-nivel 2.1. Identificar la prueba paramétrica necesaria para analizar los datos

[1] Actualmente, es común encontrar una versión más formal de la media muestral bajo la siguiente propiedad/proposición $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

Para que el estudiante identifique la prueba t-Student adecuada para analizar los datos que tiene y pueda dar solución al problema, es necesario que los siguientes indicadores estén presentes:

- Reconoce el tipo de datos que está trabajando y que provienen de poblaciones normales (e.g., número de muestras, tamaño de la muestra, son datos cualitativos o cuantitativos, estadístico muestral, parámetro poblacional y tipo de muestreo).
- Comprende el problema a resolver.
- Comprende la lógica de las pruebas t-Student para una muestra ($S1_t$) y para dos muestras ($S3_t$). Por ejemplo, la prueba t-Student para una muestra se utiliza para determinar si la media de una población, que se distribuye de acuerdo a la normal, es igual a un valor μ_0; mientras que la prueba t-Student para dos muestras busca determinar si las medias de dos poblaciones distribuidas de forma normal son iguales, en otras palabras, si no existe diferencia significativa entre las medias.
- Es capaz de seleccionar la prueba adecuada, aunque aún no puede desarrollarla.

Sub-nivel 2.2. Una aproximación a las pruebas con el estadístico t-Student

Una vez que el estudiante ha identificado la prueba, para dar respuesta al problema planteado, debe poner en juego lo siguiente:

- Es capaz de identificar la hipótesis nula que se encuentra implícita en el problema. Como se observó en la historia, lo que hoy conocemos como hipótesis nula se encontraba implícita en el problema, muchas veces en forma de pregunta, desde que surge la prueba de bondad de ajuste con el estadístico χ^2 y se extendió a otras pruebas. También se puede encontrar esta forma de hipótesis nula como primer componente en el ciclo de investigación estadística PPDAC (WILD; PFANNKUCH, 1999). Diversos estudios han reportado las bondades de este tipo de hipótesis, a través de preguntas o conjeturas (e.g., PFANNKUCH; WILD, 2004; PFANNKUCH *et al.*, 2016; STOHL; ANGOTTI; TARR, 2010).
- Conoce la distribución t-Student (e.g., es simétrica alrededor del valor cero, es más dispersa que la distribución normal y la relación grados de libertad-dispersión de la curva – conforme incrementan los grados de libertad, la dispersión de la curva t disminuye). Algunas investigaciones (e.g., BAKKER; GRAVEMEIJER, 2004; DINOV *et al.*, 2018; READING; REID, 2006; ROSSMAN, 2008;), se han apoyado con recursos tecnológicos para realizar actividades con simulaciones para promover la comprensión de la noción de distribución.

- Comprende la relación entre la distribución normal y la t-Student (e.g., a medida que los grados de libertad tienden a infinito la curva de la t se aproxima más a la normal). Puede auxiliarse de simulaciones tal como se hace para comprender la noción de distribución.

Además de los indicadores anteriores, compartidos por las pruebas t ($S1_t$ y $S3_t$), otros indicadores a considerar por prueba.

En la Prueba t-Student de la media de una muestra ($S1_t$) el estudiante puede valorar si la media de una población es igual al valor μ_0, por medio del estadístico t-Student. Para hacer esta valoración, el estudiante:

- Calcula y comprende qué indica el estadístico t-Student, utilizando conceptos/definiciones tales como muestra, población, media y desviación estándar, del $SP2_t$, y propiedades tales como $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$, correspondientes al $SP3_t$.
- Calcula y comprende los grados de libertad, de acuerdo con la propiedad del $SP3_t$ que indica que $gl = n - 1$.
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la media de la población se encuentre fuera del rango $\pm t$ (concepto/definición de probabilidad y propiedad de distribución del estadístico t-Student del $SP3_t$). Konold *et al.* (2011) realizan una crítica sobre la práctica de presentar a los estudiantes la probabilidad teórica y la experimental, e indican que, primero, es necesario fomentar un razonamiento probabilístico básico para que los estudiantes puedan realizar inferencias informales. En este sentido, consideramos que es necesario utilizar la probabilidad como una medida de certidumbre asociada a un evento previo a compararla con un límite preestablecido o un nivel de significancia.

Los elementos matemáticos que intervienen en estos indicadores pertenecen al $SP2_t$ y $SP3_t$, los cuales se caracterizaron a partir de los trabajos de Gosset (STUDENT, 1908) y Fisher (1922), principalmente.

Proceso de generalización intra-nivel

El nivel 2.2, hay un proceso de generalización en el uso del estadístico t-Student, al inicio se utiliza para una muestra y posteriormente podemos ver su uso para dos muestras que cuentan con tamaños y varianzas iguales.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

En las Pruebas t-Student para dos muestras ($S3_t$), una vez que el estudiante ha identificado que los tamaños de las muestras y las varianzas son iguales, puede valorar si las medias de dos poblaciones (distribuidas normalmente) son iguales. Para llevar a cabo esta valoración, los siguientes indicadores son necesarios:

- Comprende y es capaz de obtener la varianza y desviación estándar agrupada. Esto a partir de los objetos matemáticos primarios del $SP8_t$, tales como el concepto/definición de desviación estándar agrupada (como un método para calcular la desviación estándar agrupando las sumas de cuadrados de las dos muestras y dividiendo por el número total de grados de libertad); y la propiedad/proposición de la varianza agrupada, entendida como: $s^2 = \frac{1}{(n_1-1)(n_2-1)}\left(\sum(x_1-\bar{x}_1)^2 + \sum(x_2-\bar{x}_2)^2\right) = \frac{1}{2}\left(s_{x_1}^2 + s_{x_2}^2\right)$. También el concepto/definición (del $SP7_t$) de tamaño muestral, entendido como el número de sujetos que componen la muestra, es decir, el número de observaciones del subconjunto de la población. Cabe señalar que la fórmula de la varianza agrupada podría ser reescrita en términos actuales, considerando el uso de subíndices para poder ver el recorrido de las variables (x_1 y x_2).
- Comprende el error estándar, como la desviación estándar de la distribución de la diferencia entre las medias (concepto/definición del $SP7_t$).
- Comprende qué indica el estadístico t-Student para dos muestras, a partir de conceptos/definiciones (del $SP2_t$) como muestra, población y media; y la propiedad/proposición (del $SP8_t$) $t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{s^2\left({}^2/_n\right)}} = \frac{\bar{x}_1-\bar{x}_2}{s\sqrt{{}^2/_n}}$.
- Comprende y calcula los grados de libertad como $gl = 2n - 2$ (propiedad del $SP8_t$).
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la diferencia de las medias se encuentre fuera del rango $\pm t$ (concepto del $SP7_t$).

4.3 Nivel 3

Los indicadores que se encuentran en este nivel se pueden considerar pre-formales, pero con un mayor grado de formalidad que en el nivel 2, dado que se tiene una visión más amplia de las posibilidades para trabajar con las pruebas t-Student. Los aspectos que marcan cierto grado preformal son, por ejemplo, la forma de plantear las hipótesis (el lenguaje) y la forma de trabajar y comprender la significancia.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

Sub-nivel 3.1. Restricciones de las pruebas t-Student

El estudiante reconoce las restricciones que se tienen al realizar pruebas de hipótesis con el estadístico t-Student.

- Comprende que puede trabajar con tamaños muestrales pequeños si las muestras provienen de distribuciones normales.
- Identifica si las muestras son dependientes o independientes y comprende sus implicaciones.
- Comprende las implicaciones cuando se tiene igualdad o desigualdad de varianzas y del tamaño de las muestras.
- Es capaz de hacer explícita la hipótesis nula y alternativa en lenguaje natural.

Además, se proponen algunos indicadores por prueba.

En las Muestras dependientes ($S3_t$), una vez que el estudiante ha identificado que se trata de muestras dependientes, puede valorar si la media de las diferencias es igual al valor μ_0. Para ello:

- Calcula y comprende el estadístico t-Student como $t = \frac{\bar{x}_D - \mu_0}{s_D/\sqrt{n}}$ (a partir de lo expresado en el nivel previo y con las propiedades/proposiciones media de la diferencia, desviaciones y estadístico t, del $SP9_t$).
- Calcula y comprende los grados de libertad como $gl = n - 1$ (propiedad del $SP3_t$).
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la media de las diferencias se encuentre fuera del rango $\pm t$ (definición de probabilidad del $SP9_t$).

Proceso de generalización inter-nivel

El nivel 2.2, que considera objetos matemáticos y procesos propios del $SP2_t$ y $SP3_t$, se puede aplicar cuando se trabaja con una muestra; en cambio, bajo los elementos del nivel 3.1 ($SP9_t$), se puede trabajar con dos muestras que sean dependientes o emparejadas.

En las muestras independientes con $\boldsymbol{n_1 \neq n_2}$ ($S3_t$), si el estudiante ya ha identificado que se trata de muestras independientes y que los tamaños de las muestras son desiguales, puede valorar si las medias de dos poblaciones son iguales:

- Comprende y es capaz de obtener la desviación estándar agrupada, de acuerdo con conceptos/definiciones (propias del $SP8_t$) tales como la desviación estándar agrupada, y la propiedad $s = \sqrt{\frac{(n_1-1)(s_{x_1}^2)+(n_2-1)(s_{x_2}^2)}{n_1+n_2-2}}$, propia del $SP7_t$.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

- Calcula y comprende el estadístico t-Student como $t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ (propiedad del $SP7_t$).
- Comprende y calcula los grados de libertad como $gl = n_1 + n_2 - 2$ (propiedad del $SP7_t$).
- Comprende y es capaz de establecer relaciones con la prueba para dos muestras del nivel 2.3, como las relaciones del estadístico t-Student, los grados de libertad y la desviación estándar agrupada.

Proceso de generalización inter-nivel

En el nivel 2.3 se trabaja con el estadístico t-Student ($SP8_t$), los grados de libertad y la desviación estándar agrupada, sólo cuando $\mathrm{n}_1 = \mathrm{n}_2$. Mientras que en el nivel 3.1 se puede trabajar sin la restricción de la igualdad de los tamaños muestrales, es decir cuando $\mathrm{n}_1 \neq \mathrm{n}_2$ ($SP7_t$).

En las muestras independientes con $\boldsymbol{n_1} \neq \boldsymbol{n_2}$ y $\boldsymbol{\sigma_1^2} \neq \boldsymbol{\sigma_2^2}$ o $\boldsymbol{n_1} = \boldsymbol{n_2}$ y $\boldsymbol{\sigma_1^2} \neq \boldsymbol{\sigma_2^2}$ ($S3_t$), el estudiante ha sido capaz de identificar que está trabajando con dos muestras que tienen diferentes tamaños y diferentes varianzas; ahora, para valorar si las medias de dos poblaciones son iguales es necesario que el estudiante refleje en su práctica los siguientes indicadores:

- Es capaz de estimar la varianza de las dos poblaciones por medio de las muestras y comprende por qué no se utiliza la varianza agrupada, como en las otras pruebas t-Student para dos muestras. Lo anterior a partir del uso, principalmente, de conceptos/definiciones tales como desviación estándar ($SP2_t$) y propiedades/proposiciones (del $SP12_t$) tales como la varianza estimada y varianza de la muestra $\sigma^2 = \frac{1}{n_1}s_1^2 + \frac{1}{n_2}s_2^2$, $s_i^2 = \frac{1}{n_i - 1}\sum(x_a - \bar{x})^2$.
- Calcula y comprende el estadístico t-Student, para dos muestras independientes con desigualdad en las varianzas poblacionales, como $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1}s_1^2 + \frac{1}{n_2}s_2^2}}$ (propiedad/proposición del $SP12_t$).
- Comprende y calcula los grados de libertad como $gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$ (propiedad/proposición del $SP12_t$).
- Comprende que este método también lo puede aplicar cuando se tienen tamaños muestrales iguales.

Los elementos matemáticos que pertenecen al $SP12_t$ se caracterizaron a partir de los trabajos de Welch (1947), en los cuales realizó una adaptación para la prueba t-Student para

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

cuando las varianzas poblacionales son desiguales.

En las muestras independientes con coeficientes de regresión ($S3_t$), el estudiante puede comparar los coeficientes de regresión cuando las series de la variable independiente no son idénticas y valorar la diferencia de los coeficientes de regresión:

- Puede obtener el coeficiente b y comprende su significado, a partir del uso de conceptos/definiciones tales como variable independiente, variable dependiente y regresión, y las propiedades/proposiciones función de regresión $Y = a + b(x - \bar{x})$ y $b = \frac{S\{y(x-\bar{x})\}}{S(x-\bar{x})^2}$ (objetos matemáticos primarios propios del $SP10_t$).
- Calcula y comprende qué indica el estadístico t-Student de acuerdo con $t = \frac{b\prime - b}{\sqrt{(s^2/S_A(x-\bar{x})^2) + (s^2/S_B(x-\bar{x})^2)}}$ (propiedad/proposición del $SP11_t$).
- Comprende y calcula los grados de libertad como $gl = n_A + n_B - 2$ (propiedad/proposición del $SP11_t$).
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la diferencia de los coeficientes de regresión se encuentre fuera del rango $\pm t$. En un sentido similar a lo expresado en las otras pruebas para dos muestras.
- Reconoce que esta prueba sigue la misma lógica de la prueba de diferencia de medias de dos muestras.

Sub-nivel 3.2. Conexiones y argumentos

En este subnivel el estudiante trabaja con la significancia en una etapa previa a lo formal, así como se trabajó en las primeras décadas del siglo XX, antes de la metodología de inferencia. También, hemos adaptado de la historia la forma en que el estudiante puede rechazar o no la hipótesis nula, en las décadas mencionadas no hablaban de rechazar una hipótesis propiamente, pero si hacían la relación entre la pregunta o suposición (Ho) y el límite como desviación significativa para dar respuesta al problema.

- Comprende la significancia como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria (concepto/definición del SP2t).
- Puede encontrar el valor del estadístico teórico, en la tabla de probabilidad de la distribución t-Student, con respecto a cierta P y n; y lo contrasta con el valor t calculado. Esta era una práctica que realizaba Fisher (1934) en algunas pruebas t-Student.
- Es capaz de rechazar o no rechazar la hipótesis nula (y comprende lo que significa) bajo un contraste con un límite preestablecido como desviación significativa – esto a partir de la regla de decisión con respecto al contraste del estadístico teórico y del calculado

(propiedad del $SP3_t$), y la regla de decisión con respecto al contraste de la probabilidad obtenida a partir del estadístico calculado y un límite preestablecido (propiedad del $SP4_t$).

- Logra argumentar, con base en la significancia, por qué acepta o rechaza la hipótesis nula.
- Es capaz de conectar los resultados de la prueba con el contexto del problema. Es necesario que se fomenten espacios para que los estudiantes puedan elaborar conclusiones en términos del problema inicial, y que proporcionen argumentos con base en los resultados de las pruebas, ya que de acuerdo con los ciclos de investigación estadística es necesario que, una vez analizados los datos, estos puedan ser interpretados para elaborar conclusiones, como es el caso de lo expuesto en la dimensión uno denominada el ciclo de la investigación (PPDAC) del modelo propuesto por Wild y Pfannkuch (1999), como parte de la teoría de pensamiento estadístico.

Como hemos podido observar, en los indicadores de los niveles dos y tres se encuentran inmersos los tres principios clave (generalización, uso de los datos como evidencia y el empleo de lenguaje probabilístico) del marco de inferencia informal de Makar y Rubin (2009), aunque en diferente profundidad.

4.4 Nivel 4

Los indicadores que se presentan en este nivel corresponden a un razonamiento inferencial formal. Se espera que el estudiante pueda tomar decisiones basado en las técnicas estadísticas de la metodología de las pruebas de hipótesis.

Sub-nivel 4.1. Criterio para la toma de decisión

Valor-p:

- Puede plantear las hipótesis nula y alternativa con elementos lingüísticos simbólicos (e.g., $Ho: \mu = \mu_0$ y $Ha: \mu \neq \mu_0$, $Ho: \mu = \mu_0$ y $Ha: \mu > \mu_0$, y $Ho: \mu = \mu_0$ y $Ha: \mu < \mu_0$).
- Conoce y comprende los valores comunes del nivel de significancia, por ejemplo, 0.05 es muy común y se popularizó en las pruebas que realizó Fisher; también se utilizan 0.10 y 0.01 como nivel de significancia. La elección del nivel de significancia depende de la magnitud del error que se desea asumir.
- Comprende la relación entre el nivel de significancia (α) y el nivel de confianza $(1-\alpha)$.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

- Conoce, comprende y es capaz de aplicar la regla de decisión, si el $valor - p < \alpha$ se rechaza H_0, (propiedad/proposición del $SP8_t$).

Como se observó en la historia, y siguiendo con lo trabajado en los niveles, hemos abordado la prueba de hipótesis primero con el criterio *valor-p* para la toma de decisión y posteriormente con el estadístico teórico.

Valor crítico:

- Es capaz de identificar el valor teórico del estadístico t-Student, de acuerdo con el nivel de significancia y los grados de libertad.
- Puede representar gráficamente las regiones de aceptación y de rechazo.
- Conoce, comprende y es capaz de aplicar la regla de decisión, si t se encuentra en región de rechazo (RR) se rechaza H_0 (e.g., RR de dos colas: $t \geq t_{\alpha/2,gl}$ o $t \leq -t_{\alpha/2,gl}$; RR cola izquierda: $t \leq -t_{\alpha,gl}$; y RR cola derecha: $t \geq t_{\alpha,gl}$).
- Es capaz de brindar una respuesta al problema utilizando los resultados de la prueba y realizar procesos de argumentos con fundamentación estadística.

Sub-nivel 4.2. Error tipo I y II, y Potencia de la prueba.

- Comprende cuándo se comete el Error tipo I y la probabilidad de cometerlo. $P\begin{bmatrix} Rechazar\ H_0 | H_0 \\ es\ verdadera \end{bmatrix} = \alpha$
- Comprende cuándo se comete el Error tipo II y la probabilidad de cometerlo. También reconoce que existe un valor de β para cada posible valor del parámetro en la hipótesis alternativa, comprendiendo que se trata de una función. $P\begin{bmatrix} No\ rechazar\ H_0 | H_0 \\ es\ falsa \end{bmatrix} = \beta$
- Es capaz de calcular la probabilidad de tomar la decisión correcta cuando H_0 es verdadera y cuando H_0 es falsa.
- Comprende las relaciones entre los dos tipos de error. Por ejemplo, puede auxiliarse de una representación gráfica y reconocer en ella que cuanto mayor es α, menor es β.
- Comprende la potencia de la prueba y es capaz de calcularla. Por ejemplo, reconoce que la potencia es una función que para cada posible valor del parámetro le hace corresponder la probabilidad de que la hipótesis resulte rechazada. También que la potencia está relacionada con el tamaño de la muestra y el nivel de significancia, y que cuanto más lejana está H_a de H_0 la probabilidad de cometer el error tipo II es menor, y por lo tanto la potencia de la prueba es mayor. $P\begin{bmatrix} Decidir\ H_1 | H_1 \\ es\ verdadera \end{bmatrix} = 1 - \beta$
- Puede argumentar sobre la validez de la inferencia realizada.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

5 Prácticas Asociadas a los Niveles de RI propuestos

En esta sección presentamos ejemplos de prácticas que *activan* cada nivel. Cabe señalar que otros análisis, más detallados pueden desarrollarse, no pretendemos ser exhaustivos en las posibles respuestas y sus análisis, sólo damos una mirada del uso de la propuesta de niveles. Para ello, hemos adaptado un problema de Fisher (1934) el cual nos permitirá transitar por los cuatro niveles de RI propuestos para el estadístico t-Student.

Situación-Problema

A continuación, se presentan los datos de una muestra que fueron recabados durante un experimento de un fármaco para dormir. Los datos de la Tabla 1 son sobre el aumento de horas de sueño cuando han ingerido el fármaco. ¿Existe realmente un aumento en las horas de sueño al utilizar el fármaco?

Tabla 1 – Horas adicionales de sueño ganadas por el uso del fármaco

Paciente	Horas adicionales
1	$+0.7$
2	-1.6
3	-0.2
4	-1.2
5	-0.1
6	$+3.4$
7	$+3.7$
8	$+0.8$
9	0.0
10	$+2.0$

Fuente: elaborado por el autor

5.1 Prácticas asociadas al Nivel 1

A partir de los datos de la tabla, podemos calcular los cuartiles y tenemos que $Q1 = -0.175$, $Q2 = 0.35$, $Q3 = 1.7$ y $Q4 = 3.7$, y elaborar los gráficos con los cuartiles (Figura 4).

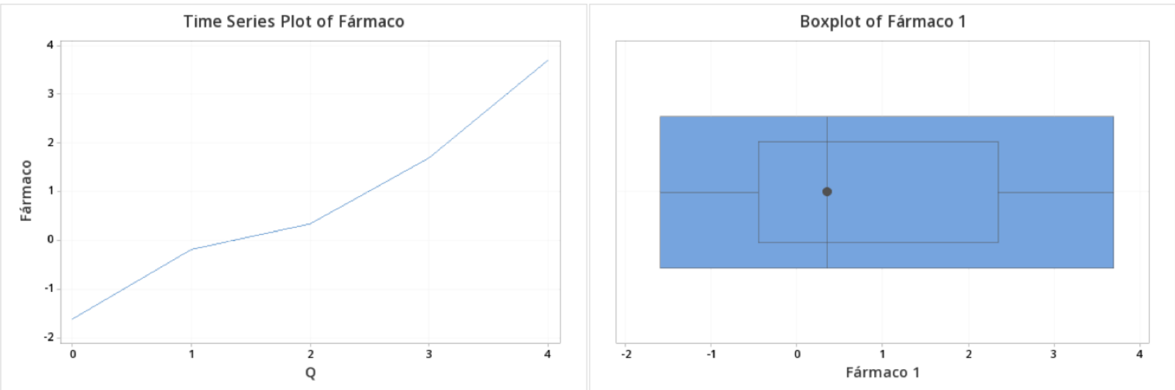

Figura 4 – Gráficos que permiten visualizar la dispersión
Fuente: elaborado por el autor con el *software* Minitab®

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

Ahora, podemos observar los cuartiles en las gráficas y vemos claramente que hay dispersión en el tamaño de los cuartiles, también podemos calcular cuánto es la amplitud de cada cuartil y tenemos que es 1.425, 0.525, 1.35 y 2. Así, podemos indicar que la serie de datos no es simétrica porque $Q2 - Q1 = 0.525$ y $Q3 - Q2 = 1.35$.

Consideramos que los análisis realizados hasta el momento no nos permiten asegurar si con el fármaco efectivamente se pueden incrementar las horas que se duerme, aunque algunos de los pacientes experimentaron una ganancia en las horas de sueño, otros pacientes incluso disminuyeron, la mediana (Q2) es cercana a cero y, en este caso, el cero indica que no hay un aumento en las horas de sueño.

Ahora, modificando la situación inicial, la Figura 5 presenta de manera gráfica la experimentación con dos fármacos ¿Existe diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2?

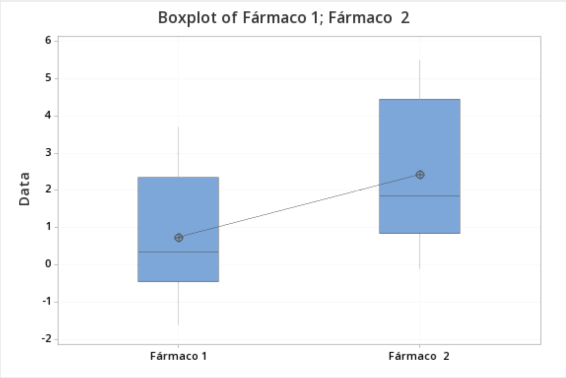

Figura 5 – Horas ganadas de sueño por fármaco
Fuente: elaborado por el autor con el *software* Minitab

A partir de la información que se presenta en la gráfica de la Figura 5, podemos decir que pareciera existir una diferencia entre las horas de sueño que se ganan al utilizar estos fármacos, aunque únicamente el último cuartil de la muestra a la que se le suministró el fármaco 2 no se superpone al diagrama de la muestra del fármaco 1, habría que tener más información sobre las dos muestras (e.g., tamaño de las muestras y tipo de muestreo) para emitir una conclusión hacía la población, pero podríamos conjeturar que el fármaco dos parece ser el más efectivo. En el diagrama podemos observar que ninguna de las dos series es simétrica, y se hace evidente una dispersión de los datos amplia en la segunda caja para ambas muestras, también que la mediana para el fármaco dos es mayor que para el fármaco 1.

Continuando con la variación de la situación (para dos fármacos), al introducir los datos de las dos muestras (Tabla 2), una respuesta de nivel 1 sería como la siguiente.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

Tabla 2 – Horas ganadas de sueño por participante con ambos fármacos

Paciente	Fármaco 1	Fármaco 2
1	$+0.7$	$+1.9$
2	-1.6	$+1.8$
3	-0.2	$+1.1$
4	-1.2	$+0.1$
5	-0.1	-0.1
6	$+3.4$	$+4.4$
7	$+3.7$	$+5.5$
8	$+0.8$	$+1.6$
9	0.0	$+4.6$
10	$+2.0$	$+3.4$

Fuente: elaborado por el autor

A partir de los datos de las Tablas 1 y 2 podemos calcular la fluctuación, para hacerlo elaboramos la Tabla 3:

Tabla 3 – Fluctuación de las horas ganadas de sueño del fármaco 2 con respecto al fármaco 1

	Fármaco 1	Fármaco 2	Diferencias	e	e^2
1	$+0.7$	$+1.9$	$+1.2$	-0.48	$+0.2304$
2	-1.6	$+1.8$	$+3.4$	$+1.72$	$+2.9584$
3	-0.2	$+1.1$	$+1.3$	-0.38	$+0.1444$
4	-1.2	$+0.1$	$+1.3$	-0.38	$+0.1444$
5	-0.1	-0.1	0.0	-1.68	$+2.8224$
6	$+3.4$	$+4.4$	$+1.0$	-0.68	$+0.4624$
7	$+3.7$	$+5.5$	$+1.8$	$+0.12$	$+0.0144$
8	$+0.8$	$+1.6$	$+0.8$	-0.88	$+0.7744$
9	0.0	$+4.6$	$+4.6$	$+2.92$	$+8.5264$
10	$+2.0$	$+3.4$	$+1.4$	-0.28	$+0.0784$
Sumatoria	$+7.5$	$+24.3$	$+16.8$	$+3.10\,E-15$	$+16.156$
Media	$+0.75$	$+2.43$	$+1.68$		$+\mathbf{1.6156}$

Fuente: elaborado por el autor

Así, obtenemos que 1.6156 es el límite de la fluctuación y $F = 1.9233$, por lo tanto, la fluctuación (F) excede el límite calculado, es decir, hay una variación de la serie de datos del fármaco 2 con respecto a la serie del fármaco 1. Esto nos indica que existe una diferencia entre las horas ganadas de sueño que manifestaron los pacientes de ambas muestras. Lo anterior puede llevar a pensar que existe diferencia entre la eficacia de los dos fármacos.

5.2 Prácticas asociadas al Nivel 2

Retomando la situación inicial, como es una muestra de tamaño pequeño, que proviene de una población normal, desconocemos el valor de la varianza poblacional y queremos saber si existe un incremento en las horas de sueño (la cual parece ser la hipótesis que se tiene), podemos utilizar una prueba t-Student para una muestra, la cual nos ayudará a determinar si la media de la población es igual a un valor μ_0, en este caso $\mu_0 = 0$.

Si aplicamos la prueba t-Student para una muestra, primero calculamos el valor del

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

estadístico bajo la expresión $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{0.75-0}{1.79/\sqrt{10}} = \frac{0.75}{0.566} = 1.325$

Ahora, procedemos a calcular los grados de libertad bajo la propiedad $gl = n - 1 = 9$. A partir del valor del estadístico t-Student y de los grados de libertad, buscamos la probabilidad en una tabla de distribución de probabilidad t-Student, y resulta que para un valor del estadístico de 1.3 con 9 gl, la probabilidad es de 0.113, es decir, que podríamos ver 887 veces en 1000 experimentos, datos como estos cuando la media de la población es igual a 0. Esto quiere decir que no tenemos evidencia suficiente para concluir que con el fármaco se incrementan las horas de sueño.

Ahora, sí retomamos la variante de la situación (dos muestras), cuyos datos se presentaron en la Tabla 2, una práctica asociada al nivel 2 sería: tenemos dos muestras pequeñas, ambas siguen de una distribución normal y desconocemos las varianzas poblacionales, como no tenemos mayor información asumiremos que son iguales, esto debido a que la varianza de la muestra del fármaco 1 es similar a la varianza de la muestra del fármaco 2. La situación cuestiona sobre si existe una diferencia entre las horas de sueño que se ganan con los fármacos 1 y 2, entonces tomemos esta duda sobre la existencia de una diferencia como la hipótesis. Con el tipo de datos que tiene el problema podemos ver que es factible utilizar una prueba t-Student para dos muestras, esta prueba busca determinar si no existe diferencia significativa entre las medias. Para llevarla a cabo es necesario calcular el valor del estadístico t-Student, determinar los grados de libertad y encontrar la probabilidad en la tabla de la distribución t.

$$\frac{\bar{x}_1-\bar{x}_2}{s\sqrt{2/n}} = \frac{0.75-2.43}{1.8667\sqrt{2/10}} = -2.01$$

La desviación estándar se obtuvo agrupando las sumas de cuadrados de las dos muestras y dividiendo por el número total de grados de libertad, es decir, calculamos la desviación estándar agrupada. Mientras que, los grados de libertad, los determinamos a partir de la formula $gl = 2n - 2 = 18$. Así, tenemos una probabilidad de las tablas de 0.06 ($0.03 + 0.03$ de cada cola), esto quiere decir que podríamos ver datos como los de estas muestras únicamente en 6 ocasiones de 100 experimentos si no existiera diferencia entre las medias de las poblaciones. A partir de la prueba que hemos realizado podemos decir que existe una diferencia entre la media de horas de sueño que se ganan con el fármaco 1 y con el 2.

5.3 Prácticas asociadas al Nivel 3

Consideremos la variación de la situación (Tabla 2), pero ahora sabemos que las

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

muestras no son independientes. Una práctica asociada a este nivel podría considerar la práctica asociada al nivel 2 y lo siguiente: Como estamos trabajando con dos muestras dependientes y con un tamaño muestral pequeño, utilizaremos una prueba t-Student para dos muestras dependientes. Tenemos como hipótesis nula que no existe diferencia significativa entre la media de las diferencias y μ_0; y como hipótesis alternativa que sí, existe una diferencia significativa entre la media de las diferencias y μ_0. $t = \frac{\bar{x}_D - \mu_0}{s_D/\sqrt{n}} = \frac{-1.68-0}{1.34/\sqrt{10}} = -3.9646$

En este caso $\mu_0 = 0$, y tanto la media como la desviación estándar que utilizamos en el estadístico no corresponden a una o ambas muestras, sino a la de la diferencia entre los datos de las dos muestras. Los grados de libertad son $gl = n - 1 = 9$, y en la tabla de probabilidad de la distribución t-Student encontramos que para ese valor del estadístico (± 3.9646, porque como se trata de una diferencia se evalúa en dos colas) se tiene una probabilidad de 0.004. Si consideramos $P = 0.05$ como límite de desviación significativa, tenemos que el valor de la probabilidad obtenida es menor que el límite, por lo que tendríamos que decir que sí, existe una diferencia significativa. También podemos valorar las hipótesis, bajo este mismo límite, con el estadístico teórico $t_{.05/2,9} = \pm 2.262$ y si lo comparamos con el estadístico calculado, tenemos que es más grande el estadístico calculado. Esto quiere decir que sobrepasa nuestro límite, por lo que las desviaciones de las expectativas son claramente significativas.

5.4 Prácticas asociadas al Nivel 4

Además de lo expresado en la respuesta del nivel previo, una respuesta en este nivel podría considerar los siguientes aspectos: La situación no indica el nivel de significancia, por lo que es posible utilizar el más común, un $\alpha = 0.05$.

Hipótesis nula $Ho: \mu = \mu_0$

Hipótesis alternativa $Ha: \mu > \mu_0$

Utilizando un *software* se obtiene la información de la Figura 6.

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for
10	0,750	1,789	0,566	-0,287

μ: population mean of Fármaco 1

Test

Null hypothesis	$H_0: \mu = 0$
Alternative hypothesis	$H_1: \mu > 0$

T-Value	P-Value
1,33	0,109

Figura 6 – Prueba t-Student para una muestra
Fuente: elaborado por el autor con el *software* Minitab

Dado que tenemos un valor del estadístico $t = 1.33$ y con un valor-p de 0.109, podemos ver que nuestro valor-p es mayor que el alfa, por lo que no tenemos evidencia suficiente para

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

rechazar Ho. Otro criterio para tomar la decisión es contrastar el valor del estadístico que calculamos con el teórico $t_{.05,9} = 1.833$, tenemos, entonces, que $1.33 < 1.833$, por lo tanto, la diferencia no es significativa, lo que nos lleva a no rechazar Ho.

En la Figura 7 podemos observar que en la gráfica de la izquierda se muestra con rojo la región de rechazo a partir del valor del estadístico teórico (según el nivel de significancia y los grados de libertad), mientras que, en la gráfica de la derecha, podemos ver el estadístico calculado y su probabilidad asociada, y cómo se encuentra dentro de la zona de rechazo.

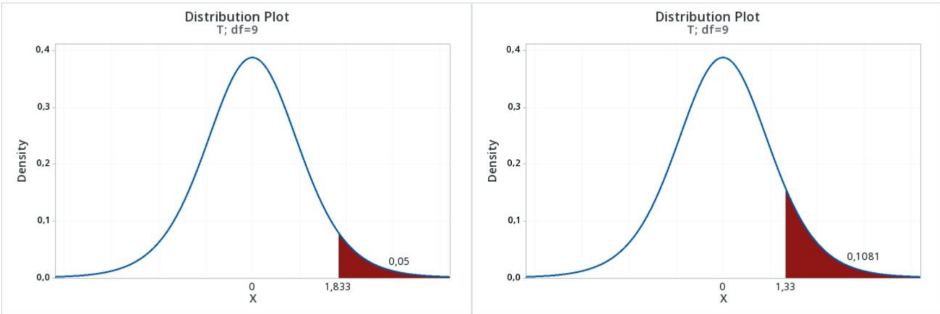

Figura 7 – Distribución t con valor del estadístico teórico (izquierda) y calculado (derecha)
Fuente: elaborado por el autor con el *software* Minitab

A partir de lo anterior, con un nivel de confianza del 95% no rechazamos Ho, es decir, no existe evidencia suficiente que nos lleve a rechazar Ho y podemos decir, con dicho nivel de confianza, que no hay diferencia significativa de la media de la población con $\mu_0 = 0$. Así, en términos del problema inicial a través de esta prueba t-Student, concluimos que no existe realmente un aumento en las horas de sueño al utilizar el fármaco 1.

En cuanto a la potencia de la prueba se tiene que es de 0.9862, entonces tenemos un muy buen valor de potencia de la prueba, esto nos indica que hay una prevalencia constante en la población, por lo que la conclusión obtenida con la prueba t-Student es muy buena. En la Figura 8 podemos ver dónde se ubica α y β, la probabilidad que tenemos de cometer el error tipo II (de no rechazar Ho cuando es falsa realmente), ya que no rechazamos Ho, es de $\beta = 0.0137$.

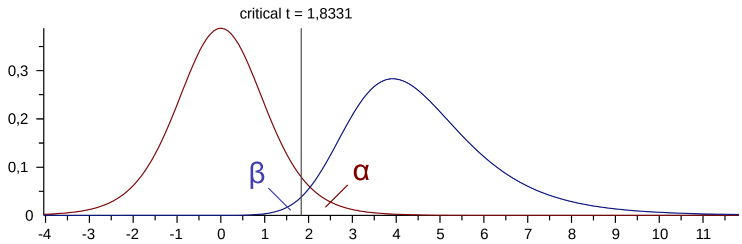

Figura 8 – Errores tipo I y tipo II para prueba t de una muestra
Fuente: elaborado por el autor con el *software* G*Power

Otro ejemplo de práctica asociada a este nivel, con una prueba t-Student con dos muestras dependientes, es la siguiente:

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

Hipótesis nula $Ho: \mu_D = \mu_0$

Hipótesis alternativa $Ha: \mu_D \neq \mu_0$

Asumimos un $\alpha = 0.05$ y utilizado un *software* para realizar la prueba t-Student se obtiene información de la Figura 9. Si aplicamos las reglas de decisión tenemos que $valor - p < \alpha$, se rechaza H_0 $0.0033 < 0.05$ y $t \geq t_{\alpha/2,gl}$ o $t \leq -t_{\alpha/2,gl}$, se rechaza H_0 $3.97 > 1.833$.

Entonces, se rechaza H_0, por lo que la diferencia no es significativa y aceptamos como probablemente cierta la Ha. Además, podemos concluir con un nivel de confianza del 95% que la media de las diferencias (de las horas de sueño ganadas por el fármaco 1 y 2) es significativamente diferente de 0.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Fármaco 1	10	0,750	1,789	0,566
Fármaco 2	10	2,430	1,941	0,614

Estimation for Paired Difference

Mean	StDev	SE Mean	95% CI for μ_difference
-1,680	1,340	0,424	(-2,638; -0,722)

μ_difference: population mean of (Fármaco 1 - Fármaco 2)

Test

Null hypothesis H_0: μ_difference = 0
Alternative hypothesis H_1: μ_difference ≠ 0

T-Value	P-Value
-3,97	0,0033

Figura 9 – Prueba t para dos muestras dependientes
Fuente: elaborado por el autor con el *software* Minitab

Ya que hemos rechazado H_0, es importante indicar que la probabilidad de haberla rechazado cuando es realmente verdadera es de 0.05. Mientras que la potencia de la prueba es de 0.9401, esta probabilidad nos indica que los resultados de la prueba son bastante confiables y que corresponde a una verdadera diferencia en la población. En la Figura 10 podemos observar los gráficos de distribución donde se ha identificado a α y β.

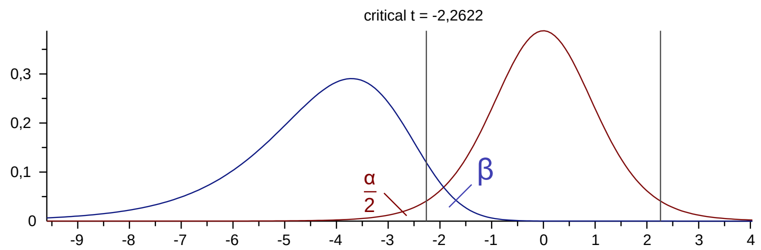

Figura 10 – Errores tipo I y tipo II para prueba t de dos muestras dependientes
Fuente: elaborado por el autor con el software G*Power

6 Reflexiones Finales

En este artículo nos propusimos presentar una propuesta de niveles progresivos de razonamiento inferencial, de lo informal a lo formal, sobre el estadístico t-Student, desde la riqueza matemática recuperada con el estudio histórico-epistemológico sobre este estadístico, y de los aportes de las investigaciones de Educación Estadística sobre el razonamiento inferencial. Nuestra propuesta consta de cuatro niveles progresivos, los indicadores del primer nivel están estrechamente vinculados con un RII, mientras que, los indicadores del segundo y

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

tercer nivel, presentan rasgos tanto de inferencia informal como de inferencia formal en distinta gradualidad, por lo que los denominamos pre-formales, y los indicadores del cuarto nivel están vinculados con el RIF.

Es importante resaltar que estos niveles se relacionan con las prácticas matemático-estadísticas que se desarrollan para resolver una situación/problema, en este caso de la t-Student; es decir, que los niveles se prevén como predictores de las prácticas y no tanto para la clasificación de problemas. Dicho de otro modo, un mismo problema/situación puede activar prácticas asociadas a cualquiera de los cuatro niveles de RI que proponemos, tal como se ha visto en los ejemplos de la sección anterior.

Los ejemplos de prácticas que *activan* cada nivel no pretenden ser exhaustivos, sino mostrar uno de los usos de esta propuesta de niveles. Estos niveles también pueden utilizarse para la planeación de clases, diseñar actividades que promuevan el razonamiento inferencial de forma progresiva sobre el estadístico t-Student, así como para realizar estudios sobre el nivel de razonamiento inferencial que se promueve con las prácticas matemáticas sobre este estadístico del currículo y los libros de texto. Además, podrían permitir caracterizar el nivel de RI sobre el estadístico t-Student, de estudiantes y profesores, aunque todo esto se prevé como líneas de investigación que se generan a partir de nuestra propuesta.

Otra línea que queda abierta para futuras investigaciones, tiene que ver con el estudio de las relaciones entre los indicadores de los niveles aquí propuestos y el currículo de matemáticas. Los indicadores que se proponen para el nivel 4 comúnmente suelen trabajarse en los primeros cursos universitarios, pero nos preguntamos, ¿en qué niveles educativos se pueden promover los indicadores de los niveles 1, 2 y 3? Nuestra propuesta inicial es que estos indicadores se podrían trabajar en la enseñanza básica y media. De manera preliminar, consideramos que los indicadores del nivel 3 podrían promoverse en tercero y cuarto medio (16-17 años), los indicadores que pertenecen al nivel 2 en primero y segundo medio (14-15 años), mientras que los del nivel 1 se podrían trabajar en el nivel básico, específicamente en sexto, séptimo y octavo básico (11-13 años) como parte de la alfabetización estadística y la promoción de inferencias informales.

La propuesta que se presenta en este artículo no pretende ser definitiva, consideramos que tanto su aplicación como otros estudios de este tipo pueden enriquecer estos niveles de razonamiento inferencial sobre el estadístico t-Student.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

Agradecimiento

Esta investigación ha sido desarrollada en el marco del Proyecto Fondecyt 1200005, financiado por la Agencia Nacional de Investigación y Desarrollo (ANID) de Chile.

Referencias

AKÉ, L. P. **Evaluación y desarrollo del razonamiento algebraico elemental en maestros en formación**. 2013. 365f. Tesis (Doctorado en Didáctica de la Matemática) – Facultad de Educación, Universidad de Granada, Granada, 2013.

BAKKER, A.; BEN-ZVI, D.; MAKAR, K. An inferentialist perspective on the coordination of actions and reasons involved in making a statistical inference. **Mathematics Education Research Journal**, Berlin, v. 29, n. 4, p. 455-470, ene. 2017.

BAKKER, A.; GRAVEMEIJER, K. Learning to reason about distribution. *In:* BEN-ZVI, D.; GARFIELD, J. (Eds.). **The challenge of developing statistical literacy, reasoning, and thinking**. Dordrecht: Kluwer Academic Publishers, 2004. p. 147-168.

BEN-ZVI, D.; ARIDOR-BERGER, K. Children's wonder how to wander between data and context. *In:* BEN-ZVI, D.; MAKAR, K. (Eds.). **The teaching and learning of statistics**: international perspectives. New York: Springer, 2016. p. 25-36.

CALLINGHAM, R.; WATSON, J. M. The Development of Statistical Literacy at School. **Statistics Education Research Journal**, Auckland, v. 16, n. 1, p. 181-201, may. 2017.

DINOV, I. D.; PALANIMALAI, S.; KHARE, A.; CHRISTOU, N. Randomization-based statistical inference: A resampling and simulation infrastructure. **Teaching Statistics**, London, v. 40, n. 2, p. 64-73, abr. 2018.

DOERR, H. M.; DELMAS, R; MAKAR, K. A modeling approach to the development of students' informal inferential reasoning. **Statistics Education Research Journal**, Auckland, v. 16, n. 2, p. 86-115, nov. 2017.

EDGEWORTH, F. Y. On methods of ascertaining variations in the rate of births, deaths, and marriages. **Journal of Statistical Society of London**, London, v. 48, n. 4, p. 628-649, dic. 1885.

FISHER, R. A. The goodness of fit of regression formulae, and the distribution of regression coefficients. **Journal of the Royal Statistical Society**, London, v. 85, n. 4, p. 597-612, jun. 1922.

FISHER, R. A. **Statistical methods for research workers**. 5. ed. Edinburgh: Oliver and Boyd, 1934.

FONT, V.; RUBIO, N. V. Procesos matemáticos en el enfoque ontosemiótico. *In:* CONTRERAS, J. M.; ARTEAGA, P.; CAÑADAS, G. R.; GEA, M. M.; GIACOMONE, B.; LÓPEZ-MARTÍN, M. M. (Eds.). **Actas del Segundo Congreso International Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos**. Granada: Universidad de Granada, 2017. p. 1-21.

GALTON, F. IV. Statistics by intercomparison, with remarks on the law of frequency of error. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science: 4a serie**, London, v. 49, n. 322, p. 33-46, 1875.

GARFIELD, J.; BEN-ZVI, D. **Developing students' statistical reasoning: Connecting research and**

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

teaching practice. New York: Springer, 2008.

GARFIELD, J.; LE, L.; ZIEFFLER, A.; BEN-ZVI, D. Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. **Educational Studies in Mathematics**, New York, v. 88, n. 3, p. 327-342, mar. 2015.

GODINO, J. D.; BATANERO, C. Significado institucional y personal de los objetos matemáticos. **Recherches en Didactique des Mathématiques**, Grenoble, v. 14, n. 3, p. 325-355, 1994.

GODINO, J. D.; BATANERO, C.; FONT, V. The onto-semiotic approach to research in mathematics education. **ZDM**, Berlin, v. 39, n. 1-2, p. 127-135, mar. 2007.

GODINO, J. D.; BATANERO, C.; FONT, V. The onto-semiotic approach: implications for the prescriptive character of didactics. **For the Learning of Mathematics**, Vancouver, v. 39, n. 1, p. 37-42, 2019.

HARRADINE, A.; BATANERO, C.; ROSSMAN, A. Students and Teachers' Knowledge of Sampling and Inference. *In:* BATANERO, C.; BURRILL, G.; READING, C. (Eds.). **Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education**. Dordrecht: Springer, 2011. p. 235-246.

JACOB, B. L.; DOERR, H. M. Statistical Reasoning with the sampling distribution. *In:* MAKAR, K.; DE SOUSA, B.; GOULD, R. (Eds.). **Sustainability in statistics education**: Proceedings of the Ninth International Conference on Teaching Statistics. Voorburg: International Statistical Institute, 2014. p. 1-6.

KONOLD, C.; MADDEN, S.; POLLATSEK, A.; PFANNKUCH, M.; WILD, C.; ZIEDINS, I.; FINZER, W.; HORTON, N. J.; KAZAK, S. Conceptual challenges in coordinating theoretical and data-centered estimates of probability. **Mathematical Thinking and Learning**, Londres, v. 13, n. 1-2, p. 68-86, ene. 2011.

MAKAR, K.; BEN-ZVI, D. The Role of Context in Developing Reasoning about Informal Statistical Inference. **Mathematical Thinking and Learning**, Londres, v. 13, n. 1-2, p. 1-4, ene. 2011.

MAKAR, K.; RUBIN, A. A framework for thinking about informal statistical inference. **Statistics Education Research Journal**, Auckland, v. 8, n. 1, p. 82–105, may. 2009.

MAKAR, K.; RUBIN, A. Learning about statistical inference. *In:* BEN-ZVI, D.; MAKAR, K.; GARFIELD, J. (Eds.). **International handbook of research in statistics education**. Switzerland: Springer International, 2018. p. 261-294.

MINEDUC. Ministerio de Educación de Chile. **Bases Curriculares 3º y 4º medio**. Santiago: Unidad de Currículum y Evaluación, 2019.

MOLINA, O. J. **Sistema de normas que influyen en procesos de argumentación**: un curso de geometría del espacio como escenario de investigación. 2019. 511 f. Tesis (Doctorado en Educación Matemática) – Escuela de Postgrado, Universidad de Los Lagos, Osorno, 2019.

NOLL, J.; HANCOCK, S. Proper and paradigmatic metonymy as a lens for characterizing student conceptions of distributions and sampling. **Educational Studies in Mathematics**, New York, v. 88, n. 3, p. 361-383, mar. 2015.

PFANNKUCH, M. Reimagining curriculum approaches. *In:* BEN-ZVI, D.; MAKAR, K.; GARFIELD, J. (Eds.). **International handbook of research in statistics education**. Switzerland: Springer International, 2018. p. 387-413.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v35n71a25

PFANNKUCH, M.; ARNOLD, P.; WILD, C. J. What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. **Educational Studies in Mathematics**, New York, v. 88, n. 3, p. 343-360, mar. 2015.

PFANNKUCH, M.; BUDGETT, S.; FEWSTER, R.; FITCH, M.; PATTENWISE, S.; WILD, C.; ZIEDINS, I. Probability modeling and thinking: What can we learn from practice? **Statistics Education Research Journal**, Auckland, v. 15, n. 2, p. 11-37, nov. 2016.

PFANNKUCH, M.; WILD, C. Towards an understanding of statistical thinking. *In:* BEN-ZVI, D.; GARFIELD, J. (Eds.). **The challenge of developing statistical literacy, reasoning and thinking**. Dordrecht: Kluwer Academic Publishers, 2004. p. 17-46.

PINO-FAN, L.; GODINO, J. D.; FONT, V. Assessing key epistemic features of didactic-mathematical knowledge of prospective teachers: the case of the derivative. **Journal of Mathematics Teacher Education**. Dordrecht, v. 21, n. 1, p. 63-94, feb. 2018.

READING, C.; REID, J. An emerging hierarchy of reasoning about distribution: From a variation perspective. **Statistics Education Research Journal**, Auckland, v. 5, n. 2, p. 46-68, nov. 2006.

ROSSMAN, A. J. Reasoning about Informal Statistical Inference: One Statistician's View. **Statistics Education Research Journal**, Auckland, v. 7, n. 2, p. 5-19, nov. 2008.

SOTOS, A. E. C.; VANHOOF, S.; VAN DEN NOORTGATE, W.; ONGHENA, P. Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. **Educational Research Review**, London, v. 2, n. 2, p. 98-113, 2007.

STOHL LEE, H.; ANGOTTI, R. L.; TARR, J. E. Making comparisons between observed data and expected outcomes: students' informal hypothesis testing with probability simulation tools. **Statistics Education Research Journal**, Auckland, v. 9, n. 1, p. 68-96, may. 2010.

STUDENT. The probable error of a mean. **Biometrika**, Oxford, v. 6, n. 1, p. 1-25, mar. 1908.

VETTEN, A. D.; SCHOONENBOOM, J.; KEIJZER, R.; OERS, B. The Development of Informal Statistical Inference Content Knowledge of Pre-service Primary School Teachers During a Teacher College Intervention. **Education Studies in Mathematics**, Amsterdam, v. 99, n. 2, p. 217-234, oct. 2018.

WELCH, B. L. The generalization of 'Student´s' problem when several different population variances are involved. **Biometrika**, Oxford, v. 34, n. 1/2, p. 28-35, ene. 1947.

WILD, C.; PFANNKUCH, M. Statistical thinking in empirical enquiry. **International Statistical Review**, Voorburg, v. 67, n. 3, p. 223-248, dec. 1999.

ZIEFFLER, A.; GARFIELD, J.; DELMAS, R.; READING, C. A framework to support research on informal inferential reasoning. **Statistics Education Research Journal**, Auckland, v. 7, n. 2, p. 40–58, nov. 2008.

Submetido em 31 de Agosto de 2020.
Aprovado em 04 de Maio de 2021.