ISSN: 2178-7727
DOI: 10.17648/acta.scientiae.5772

Nivel de razonamiento estadístico de los estudiantes chilenos de Pedagogía en Matemática sobre pruebas de hipótesis estadísticas

Chia-Shih Su [a]
Carolina Marchant [b]

[a] Universidad Católica del Maule, Programa de Doctorado de didáctica de Matemática, Talca, Chile
[b] Universidad Católica del Maule, Facultad de Ciencias Básicas, Departamento de Matemática, Física y Estadística, Talca, Chile

Recibido para publicación 29 mar. 2020. Aceptado tras revisión el 22 sep. 2021. Editora designada: Claudia Lisete Oliveira Groenwald

RESUMEN

Contexto: Existe una relación entre el progreso del país y la formación estadística de sus ciudadanos, donde el razonamiento estadístico de los profesores de matemática es crucial para incrementar esta relación. **Objetivo**: Determinar el nivel de razonamiento estadístico sobre pruebas de hipótesis de futuros profesores de matemática chilenos. **Diseño**: La metodología utilizada es de enfoque cuantitativo transeccional. Se utilizó la taxonomía SOLO (Structure of the Observed Learning Outcome) para caracterizar el nivel de conocimiento de profesores en formación sobre los principales conceptos de pruebas de hipótesis. **Entorno y participantes**: Estudiantes chilenos de V a VIII nivel de Pedagogía en Matemática que aprobaron las asignaturas de Estadística del itinerario formativo, asignaturas que incluyen el tópico de pruebas de hipótesis. **Recopilación y análisis de datos**: A través de un muestreo no probabilístico consecutivo y de un instrumento previamente validado por jueces expertos, se analizó una muestra conformada por 29 de 43 profesores en formación obteniendo una representación del 67% de la población objetivo. **Resultado**: Hubo suficiente evidencia para afirmar que el nivel de razonamiento estadístico los profesores en formación se encuentra en pre-estructural y uni-estructural de la taxonomía SOLO acerca de su conocimiento sobre las pruebas de hipótesis estadísticas. **Conclusión**: Es necesario corregir esta situación mediante herramientas remediales antes de que estos estudiantes egresen y se inserten en la labor docente.

Palabras clave: profesor chileno en formación; taxonomía SOLO; razonamiento estadístico; pruebas de hipótesis estadísticas.

Corresponding author: Carolina Marchant. Email: carolina.marchant.fuentes@gmail.com

Nivel de raciocínio estatístico de estudantes chilenos de Pedagogia em Matemática em testes de hipótese estatística

RESUMO

Contexto: Existe uma relação entre o progresso do país e a formação estatística de seus cidadãos, onde o raciocínio estatístico dos professores de matemática é crucial para aumentar essa relação. **Objetivo**: Determinar o nível de raciocínio estatístico nos testes de hipóteses de futuros professores chilenos. **Design**: A metodologia utilizada é uma abordagem quantitativa transecional. A taxonomia SOLO (Structure of the Observed Learning Outcome) foi utilizada para caracterizar o nível de conhecimento de professores chilenos em formação sobre os principais conceitos de teste de hipóteses. **Ambiente e participantes**: Alunos chilenos do nível V a VIII de Pedagogia em Matemática, aprovados nas disciplinas de Estatística do itinerário formativo, disciplinas que incluem o tópico de testes de hipóteses. **Coleta e análise de dados**: Por meio de amostragem consecutiva não probabilística e um instrumento previamente validado por juízes especialistas, foi analisada uma amostra composta por 29 dos 43 alunos em formação que concordaram em participar da pesquisa, obtendo uma representação de 67% da população alvo. Resultado: Existem evidências suficientes para afirmar que o nível de raciocínio estatístico dos professores em formação é pré-estrutural e uni-estrutural da taxonomia SOLO sobre seu conhecimento sobre os testes de hipóteses estatísticas. **Conclusão**: É necessário corrigir essa situação por meio de ferramentas corretivas antes que esses alunos se formem e entrem no trabalho de ensino.

Palavras-chave: professor chileno em formação; taxonomia SOLO; raciocínio estatístico; teste de hipótese estatística.

INTRODUCCIÓN

Batanero (2000, 2013a) señaló que al principio de la década de los ochenta comenzó el interés a nivel internacional sobre la enseñanza de la estadística descriptiva e inferencial. Según Gal (2002), la cultura estadística requería del desarrollo de dos capacidades en conjunto: (1) la capacidad interpretativa que se utiliza para evaluar críticamente la información estadística en diversos contextos, y (2) la capacidad comunicativa y argumentativa que se emplea para discutir las opiniones respecto de la información obtenida. Para que las futuras generaciones adquirieran estas dos capacidades, él consideró que era fundamental que los profesores de matemática que enseñan estadística, contaran con una formación en la disciplina que apuntara hacia la abstracción lógica, la alfabetización básica, numérica, y el razonamiento y pensamiento estadístico que les permitieran describir y generalizar los fenómenos observados en la vida cotidiana.

Watson (1997) estimó que era posible cuestionar argumentos y fundamentos para obtener una conclusión coherente, por eso, construyó un modelo compuesto de los siguientes 3 elementos: i) el conocimiento básico de los conceptos estadísticos y probabilísticos, ii) la comprensión del razonamiento y argumentos estadísticos presentados dentro de un contexto social y iii) una actitud crítica basada en la evidencia estadística.

Entonces, si existía desde el año 1997 un modelo estadístico que fomentaba el razonamiento y la argumentación estadística, y la actitud crítica, ¿por qué Batanero (2013a) señala que las clases universitarias de estadística hacen demasiado hincapié en el procedimiento y aplicaciones de fórmulas para obtener valores numéricos de un estadístico que no son realmente importantes y dejan que los estudiantes finalicen los cursos de estadística sin adquirir las competencias señaladas en los respectivos programas? Según Del Pino y Estrella (2012), Estrella (2014) e Inzunsa y Jiménez (2013), muchos profesores que imparten clases de Estadística son profesionales no docentes, por eso, cuando ejercen como profesor en la universidad no cuentan con herramientas docentes y didácticas adecuadas. Por otro lado, Batanero, Godino, Vallecillos, Green y Holmes (1994, p. 528) afirmaron:

> La principal dificultad en la enseñanza (universitaria) de esta materia se basa en que la estadística ha recibido, hasta la fecha, menos atención que otras ramas de la matemática como álgebra, aritmética o geometría. Además, la mayor parte de las investigaciones realizadas fueron llevadas a cabo por psicólogos en lugar de matemáticos o estadísticos y en situaciones experimentales en lugar de situaciones escolares.

En este contexto, los estudiantes pueden verse perjudicados, ya que sus profesores, en algunas ocasiones, entregan los contenidos estadísticos con dificultad y poca comprensión, desfavoreciendo su desarrollo del razonamiento y pensamiento estadístico.

ANTECEDENTES TEÓRICOS

A fin de cumplir con el objetivo de esta investigación, es necesario abordar las siguientes temáticas: 1) el razonamiento estadístico, 2) la taxonomía SOLO y sus niveles en contexto del proceso de pruebas de hipótesis, 3) nociones de las pruebas de hipótesis en la enseñanza, 4) niveles de razonamiento estadístico sobre pruebas de hipótesis estadísticas según la taxonomía SOLO, y 5) el concepto de las pruebas de hipótesis estadísticas y las dificultades y concepciones erróneas de los estudiantes.

Razonamiento estadístico

Según Muñoz (2015), Chervaney y sus colaboradores consideraron que el razonamiento estadístico se refiere a la capacidad y la habilidad de emplear los contenidos y conceptos estadísticos como recordar, reconocer y discernir los conceptos estadísticos y la habilidad en la resolución de problemas. Del Más utilizó una serie de palabras en actividades propuestas a través de preguntas en clases de Estadística y Probabilidad para una fácil aproximación del nivel de razonamiento estadístico de los estudiantes que se muestra en la Tabla 1.

Tabla 1

Palabras orientativas que aproxima el nivel de razonamiento. (Muñoz, 2015)

Nivel razonamiento	**Palabras identificadoras**
Alfabetización estadística	Identificar, describir, traducir, interpretar, leer
Razonamiento estadístico	¿Por qué? ¿Cómo? Explicar
Pensamiento estadístico	Aplicar, criticar, evaluar, generalizar

Por su parte, Garfield (2002, p.1) afirmó:

> Statistical reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, graphical representations, and statistical summaries. Much of statistical reasoning combines ideas about data and chance, which leads to making inferences and interpreting statistical results.

De las citas anteriores, queda claro que mediante el razonamiento estadístico sería posible obtener información contextualizada en base a los datos, resúmenes estadísticos, representaciones gráficas y modelos, por lo que se puede entender el razonamiento estadístico como la conexión entre la habilidad lógica y la habilidad argumentativa en el contexto del aprendizaje de las pruebas de hipótesis de estudiantes de Pedagogía en Matemática. A través de la habilidad lógica se puede decidir adecuadamente las unidades de análisis que conformaban una muestra de la población objetivo en estudio, apoyándose en aspectos teóricos para rechazar o no rechazar la hipótesis sobre una determinada situación y obtener una conclusión. Y a través de la habilidad argumentativa pueden explicar con fundamento los resultados obtenidos. Cabe destacar que, para utilizar adecuadamente una prueba de hipótesis, el estudiante debe emplear habilidades estadísticas como la alfabetización estadística, el razonamiento estadístico y el pensamiento estadístico. La habilidad de la alfabetización estadística corresponde a la habilidad que el estudiante utiliza para interpretar la situación y los datos extraídos de una muestra, identificando el o los parámetros de la población y formulando adecuadamente las hipótesis. La habilidad del razonamiento estadístico implica un proceso en que el estudiante lleva a cabo una prueba de hipótesis e identifica evidencia para decidir si debe rechazar o no la hipótesis nula basada en un estadístico de prueba de una distribución muestral determinada. Según Inzunsa y Jiménez (2013, p.11), el estudiante deber estar consciente de que la muestra sobre la cual se trabaja para probar la hipótesis es "sólo una de las posibles muestras que podrían ser extraídas de una población, y que, por lo tanto, existe el riesgo de cometer errores con cualquiera de las dos decisiones que tomará". La habilidad del pensamiento estadístico surge cuando el estudiante cuestiona y critica los modelos existentes para abordar mejor los problemas reales.

Taxonomía SOLO y los niveles de razonamiento

Dentro de los modelos que clasifican el nivel de razonamiento estadístico, la taxonomía SOLO ha sido utilizada en educación estadística para categorizar el desarrollo cognitivo de diversos conceptos estadísticos. Los cuales pueden ser visualizados en los trabajos desarrollados por Amaro y Sánchez (2015), Pfannkuch (2005a, 2005b), Inzunsa y Jiménez (2013), Reading y Reid (2006), Sánchez, García y Medina (2014), García y Sánchez (2013) y García y Hernández (2018). Inzunsa y Jiménez (2013) caracterizaron los niveles del razonamiento estadístico sobre pruebas de hipótesis a través de la taxonomía SOLO en una investigación realizada con estudiantes universitarios mexicanos. Debido a esto, se consideró utilizar la taxonomía SOLO para determinar el nivel de razonamiento estadístico de futuros profesores chilenos acerca de pruebas de hipótesis en esta investigación.

A diferencia de la teoría de Piaget que consideró un desarrollo cognitivo definido en términos de una estructura lógica a la par de las etapas que atravesaban los estudiantes, Biggs y Collis contemplaron las diferencias del aprendizaje e interacciones de los estudiantes con sus pares en clase y finalmente diseñaron la taxonomía SOLO en 1982. Por eso, la taxonomía representa un instrumento que ayuda a los profesores a determinar el nivel de desarrollo cognitivo de sus estudiantes mediante sus respuestas de una tarea específica en el contexto de la variación temporal en el aprendizaje, llamada decálages. Los decálages constituyen evidencias de cambios en el aprendizaje, en la actuación o en la motivación de los estudiantes y no cambios en el desarrollo cognitivo como manifiesta la teoría de Piaget.

Tabla 2

Niveles de la taxonomía y sus descriptores. (Biggs y Collis, 1982)

Niveles	Descriptores de la taxonomía SOLO
Uni-estructural	(Comprensión sobre un concepto o una aplicación concreta) El estudiante se enfoca en el uso de un aspecto relevante de una tarea planteada, el cual podría ser el empleo correcto de solo un concepto o un procedimiento. El estudiante que se encuentra en nivel uni- estructural sólo posee capacidades concretas como identificar, repetir y realizar un procedimiento sencillo.
Multi-estructural	(Comprensión limitada) El estudiante utiliza más de un aspecto relevante de una tarea planteada, como agilidad en clasificar, combinar, enumerar, describir, hacer una lista, o hacer un algoritmo, pero por no saber integrar todos los conceptos y procedimientos involucrados, no llega a la solución correcta.
Relacional	(Relación entre datos y teoría, y acción y finalidad) El estudiante es capaz de integrar todos los aspectos relevantes de una tarea planteada en una estructura coherente, el estudiante posee la capacidad de comparar, contrastar, explicar causas, analizar, relacionar y aplicar, y justificar la teoría de la cual aprende.
Abstracción Ampliada	(Nivel que trasciende lo tratado en la enseñanza) El estudiante es capaz de criticar y cuestionar el modelo convencional de una tarea planteada y tiene la capacidad de teorizar, generalizar y crear un nuevo modelo para una tarea a través de la investigación.

Según Huerta (1997), las interacciones de las respuestas de los alumnos con sus pares en clase que Biggs y Collis descubrieron se tratan de dos fenómenos. Biggs nombró el primer fenómeno como modos de funcionar y el segundo fenómeno, conocido como la taxonomía SOLO. La taxonomía SOLO consiste en "evaluar cualquier respuesta de los estudiantes como un fenómeno en sí mismo, sin que la respuesta representa necesariamente una etapa particular en el desarrollo intelectual" (p. 43), que fundamentalmente se refiere a una organización estructural del conocimiento separada en diferentes niveles de complejidad como el nivel pre-estructural, uni-estructural, multi-estructural, relacional y abstracción extendida.

En la Tabla 2, se entrega el detalle de los niveles de la taxonomía SOLO dada por Biggs y Collis (1982) con sus respectivos descriptores.

Nociones de las pruebas de hipótesis en la enseñanza

Según Inzunsa y Jiménez (2013), desde 1935, en las aulas se comenzó a utilizar el modelo integrado de los enfoques de Fisher y Neyman-Pearson en el proceso de una prueba de hipótesis. Para explicar mejor el modelo integrado de la lógica híbrida basada en los enfoques de Fisher y Neyman-Pearson se ha adaptado el siguiente ejemplo de la investigación de Leenen (2012).

Un determinado académico e investigador de cierta universidad organizó cursos dirigidos a profesionales del área de idiomas con el objetivo de potenciar sus habilidades comunicativas y desea evaluar la eficacia de dos estrategias para la enseñanza: la primera de ellas está relacionada a la educación tradicional, donde el curso de capacitación se imparte en una serie de clases presenciales. Y la segunda se refiere a educación a distancia, la cual ofrecía los mismos contenidos a través de una plataforma virtual y donde el contacto profesor-alumno se realiza exclusivamente por vía electrónica. Para este fin, el investigador diseñó un estudio en el marco de un curso de inglés basado en competencias con el fin de proporcionar herramientas a los profesionales que les ayudara a realizar un discurso fluidamente en inglés en un congreso internacional. Para esto, asignó aleatoriamente la mitad de los 50 profesionales inscriptos para el curso con la primera estrategia de enseñanza (educación tradicional) mientras que la otra mitad recibiría el curso con la segunda estrategia (educación a distancia). Tanto al inicio (pre) como al final (post) del curso, se aplicó a cada profesional cuatro pruebas para medir diferentes aspectos de su conocimiento sobre el tema del curso, y se obtuvo una puntuación global sumando los resultados en las cuatro pruebas. Finalmente, se obtuvo la diferencia entre pre y post de dichas puntuaciones globales para cada participante. Posteriormente se planteó una prueba de hipótesis que incluye los siguientes pasos:

1. Definición de variables y supuestos: Se trata de un conjunto de supuestos sobre las variables de interés. En este ejemplo hay dos variables, estas son:

(a) X: diferencia entre los puntajes promedios de las pruebas de pre y post de los profesionales que recibieron el curso tradicional, e

(b) Y: diferencia entre los puntajes promedios de las pruebas de pre y post de los profesionales del curso impartido a distancia.

El investigador asumió que X e Y se distribuyen normalmente con media μ_X y μ_Y, y varianza σ_X^2 y σ_Y^2, respectivamente, suponiendo que las observaciones de la muestra fueron extraídas de forma aleatoria de sus respectivas poblaciones. Los parámetros del modelo fueron μ_X, μ_Y y σ^2.

2. Formulación de hipótesis: En este contexto se formuló la prueba de hipótesis como una de diferencia entre las medias de dos poblaciones. Es decir, las hipótesis nula y alternativa fueron expresadas como de la siguiente forma:

H_0: μ_X - μ_Y = 0 vs H_1: μ_X - μ_Y $\neq 0$.

3. Definición del estadístico de prueba: En este caso, el estadístico de prueba bajo H_0 se define como:

$T = \frac{(\overline{X}-\overline{Y})-0}{\sqrt{\frac{S_X^2+S_Y^2}{n}}}$, donde $\overline{X}$ y $\overline{Y}$ corresponden a las medias muestrales de ambos grupos, S_X^2 y S_Y^2 las varianzas muestrales, y n es el número de observaciones en cada grupo, que en este caso es igual a 25 profesionales.

4. Identificación de la distribución del estadístico de prueba bajo los supuestos del modelo: El estadístico de prueba T entregado en el ítem se distribuye según la distribución t de Student con 48 grados de libertad, obtenidos desde 2n - 2.

5. Obtención del valor del estadístico de prueba según la muestra observada: Desde los 2 grupos, el investigador obtuvo que la media en el grupo que asistió a las clases presenciales fue de $\bar{x}$ = 13 puntos y en el grupo del curso a distancia $\bar{y}$ = 9 puntos, y las varianzas observadas fueron s_x^2 = 30 y s_y^2 = 45. Reemplazando la información muestral, se obtiene

$$t_{obs} = \frac{(\bar{x}-\bar{y})-0}{\sqrt{\frac{s_x^2+s_y^2}{n}}} = \frac{(13-9)-0}{\sqrt{\frac{30+45}{2}}} = 2{,}309.$$

6. Obtención del valor-*p*:

Según Montgomery y Runger (2012, p.312), "el valor p es la probabilidad de que el estadístico de prueba tome un valor que es al menos tan extremo como el valor observado del estadístico cuando la hipótesis nula H0 es verdadera, por eso, el valor p es el nivel de significación más bajo que llevaría al rechazo de la hipótesis nula H_0 con los datos dados."

Según la definición anterior, el valor-*p* se refiere a la probabilidad de observar t_{obs} o un valor más extremo en la distribución de referencia. En este ejemplo, el investigador consideró todos los valores mayores a 2,309 y menores a -2,309 más extremos que el valor del estadístico de prueba observado en la distribución t de Student con 48 grados de libertad, debido a que corresponde a un ensayo bilateral. Por lo que, el valor-p o la probabilidad de observar un valor más extremo que t_{obs} es 0,03.

7. Decisión de rechazar o no la hipótesis nula: si el valor-*p* es menor al nivel de significación α, se rechaza la hipótesis nula, en el caso contrario, no se rechaza. En este caso, el investigador realizó una comparación entre el valor-*p*=0.03 con el nivel de significación α=0,05, resultando la relación valor-*p*< α, indicando que debía rechazar la hipótesis nula en favor de la hipótesis alternativa, concluyendo que hubo suficiente evidencia muestral para afirmar que existió diferencia entre los puntajes promedio de las pruebas entre pre y post test en la educación tradicional y educación a distancia.

Según Leenen (2012), el valor-*p* se desarrolla dentro del marco frecuentista (o clásico) y que los parámetros del modelo estadístico son considerados como un valor determinado y fijo. Es decir, en las diferentes repeticiones, los parámetros tienen "el mismo valor, pero los estadísticos muestrales varían", por lo que se utiliza la distribución del estadístico de prueba para describir dicha variación en las diferentes repeticiones del experimento. Cuando el valor-*p* se interpreta como la proporción de veces en las infinitas repeticiones conceptuales y el estadístico de prueba tiene un valor tan extremo o más extremo que el valor observado en la ejecución del experimento, entonces ocurre la interpretación de valor-*p* < α, que equivale a decir que el resultado observado es inusual o que la hipótesis nula no es correcta.

Niveles de razonamiento estadístico sobre pruebas de hipótesis estadísticas según la taxonomía SOLO

De acuerdo a la descripción general de la taxonomía SOLO y la noción sobre una prueba de hipótesis estadística explicada en las secciones anteriores, se diseñaron los descriptores de los niveles de la taxonomía SOLO en términos de los conceptos de las pruebas de hipótesis estadísticas con el objetivo de facilitar la caracterización del nivel de razonamiento estadístico de los estudiantes según las respuestas entregadas en el instrumento, el cual fue diseñado y validado para ese fin. A continuación, adaptada de la definición que Biggs y Collis (1982) se presentan los descriptores de los niveles de aprendizaje de la taxonomía SOLO en función de las pruebas de hipótesis estadísticas.

1. **Pre-estructural:** en las respuestas de una tarea sólo se observa el uso aislado y superficial de los conceptos de dicha prueba de hipótesis, justificación no concordante, con muchos errores conceptuales y procedimentales, o sea, que la respuesta no cuenta con ningún aspecto relevante sobre pruebas de hipótesis o bien, podría dejar la actividad sin resolver.
2. **Uni-estructural:** en las respuestas de una tarea se observa el uso correcto de solo un aspecto relevante de la tarea planteada, evidenciando capacidades concretas como identificar, repetir y realizar un procedimiento sencillo, y en la tarea de mayor nivel, no articula los conceptos y procedimientos involucrados para poder tomar una correcta decisión sobre las hipótesis formuladas y justificarlas con la teoría apropiada.
3. **Multi-estructural:** en las respuestas de una tarea se observa el uso de más de un aspecto relevante, evidenciando la capacidad de clasificar, combinar, enumerar, describir, hacer una lista, combinar y hacer algoritmos, no obstante, al no poder integrar todos los conceptos y procedimientos involucrados, el estudiante sólo realiza una conclusión incompleta en contexto de la pregunta solicitada, por eso, no se puede decir que el estudiante tiene la comprensión total sobre el concepto y uso adecuado de las pruebas de hipótesis.
4. **Relacional:** en las respuestas de una tarea se observa una conexión entre los conceptos y procedimientos involucrados, evidenciando la capacidad de comparar, contrastar, explicar causas, analizar, aplicar, relacionar, justificar su respuesta apoyado de la teoría aprendida sobre las pruebas de hipótesis y concluir en el contexto de la pregunta solicitada.

Dificultad y concepciones erróneas sobre las pruebas de hipótesis

Una comprensión sólida de la estadística inferencial es fundamental para el diseño y la interpretación de fenómenos de la vida cotidiana y de los resultados de investigaciones en cualquier disciplina científica (Batanero, Vera y Díaz, 2012; Castro, Vanhoof, Van Den Noortgate y Onghena, 2007).

Sin embargo, Batanero (2005) observa que muchos estudiantes, incluso a nivel universitario, a menudo carecen de la capacidad de integrar diferentes ideas y usar correctamente los conceptos en el razonamiento inferencial y "alerta que (…) tienen concepciones incorrectas o son incapaces de hacer una adecuada interpretación de los resultados estadísticos" (Batanero, 2013a, p.55). Particularmente, en el tópico de pruebas de hipótesis, investigaciones como las de Batanero et al. (1994), Vallecillos (1996), Vallecillos y Batanero (1997), Castro et al. (2007), Batanero et al. (2012) e Inzunsa y Jiménez (2013) comentan que los estudiantes suelen cometer error y tener confusión en:

1. La distinción entre la hipótesis lógica e hipótesis estadística: la hipótesis lógica determina su verdad por el proceso deductivo, por eso, es falsa o verdadera siempre, en cambio, una hipótesis estadística se determina con evidencias en base a los datos de una muestra aleatoria y está sujeta a un nivel de significación, indicando que aunque se determine con evidencias que era verdadera, existe una probabilidad de que en realidad sea falsa (Vallecillos, 1996; Batanero, 2013; Inzunsa y Jiménez, 2013).
2. La formulación de las hipótesis estadísticas: la hipótesis estadística se establecen en base a los parámetros poblacionales, sin embargo, es frecuente que los estudiantes la construyan con los estadísticos muestrales. (Vallecillos, 1996; Vallecillos y Batanero, 1997; Inzunsa y Jiménez, 2013).
3. La decisión de rechazar la hipótesis nula H_0 cuando el resultado es estadísticamente significativo.
4. La definición del nivel de significación: en este apartado los estudiantes pueden cometer un error al intercambiar el suceso condición y condicionado, interpretando la definición de nivel de significación como la del error tipo I, es decir en vez de considerar el nivel de significación como $\alpha = P(\text{Rechazar } H_0 / H_0 \text{ cierta})$ lo toma como $\alpha = P(H_0 \text{ cierta /se ha rechazado } H_0)$ (Batanero et al., 1994; Vallecillos, 1996; Vallecillos y Batanero, 1997; Castro et al., 2007; Batanero et al., 2012; Inzunsa y Jiménez, 2013) .
5. El nivel de significación es determinante de la región crítica y de aceptación que influye en el criterio de decisión (Vallecillos y Batanero, 1997).
6. La distinción entre la probabilidad de cometer errores tipo I y tipo II, y entre probabilidad de error tipo II y la definición de potencia: los estudiantes a menudo interpretan la probabilidad de cometer errores de tipo I (α) y II (β) como probabilidad de sucesos

complementarios, y además tienen la confusión entre la probabilidad de error tipo I y la potencia de la prueba (Vallecillos, 1996; Batanero, 2013b).

7. Poca consideración sobre el tamaño de la muestra: los estudiantes creen poder rechazar H_0 al obtener un resultado estadísticamente significativo sin considerar el tamaño de muestra (Vallecillos, 1996).
8. Dificultad en discernir el tipo de distribución muestral a utilizar en la prueba de hipótesis (Vallecillos, 1996).

Castro et al. (2007), Batanero (2013b) e Inzunsa y Jiménez (2013) manifiestan que esos errores y dificultades que tienen los estudiantes en el tópico en estudio son transmitidos por los docentes e incluso de libros de texto. Por otro lado, investigaciones de Vallecillos (1996), Batanero et al. (2012), Castro et al. (2007) e Inzunsa y Jiménez (2013) muestran que hasta los profesionales estadísticos y docentes de matemática tienen confusión y errores conceptuales e interpretan mal los resultados obtenidos en las investigaciones.

Los resultados de estas investigaciones junto a los que fueron mencionados en la introducción de este trabajo son indicadores que resaltan la importancia del presente estudio, que es determinar el nivel de razonamiento estadístico y los conceptos erróneos que tienen los profesores de matemática en formación, para poder corregir y mejorarlos antes de que ellos se inserten en el campo laboral.

METODOLOGÍA

La metodología utilizada en este estudio fue de paradigma positivista, o bien, cuantitativa transeccional mediante aplicación de un instrumento a individuos de una muestra no probabilística consecutiva de la población objetivo. Dicha población corresponde a todos los estudiantes regulares que aprobaron la asignatura que contempla entre sus tópicos las pruebas de hipótesis estadísticas. De los 191 estudiantes regulares de la carrera, sólo 43 estudiantes en los niveles V, VI, VII y VIII de la carrera de Pedagogía en Matemática cumplieron con las condiciones de selección, constituyendo la población objetivo de este estudio. Se utilizó un muestreo consecutivo, es decir, aplicar el instrumento a todos ellos, obteniendo respuestas del instrumento de sólo 29 estudiantes, que conformaron una muestra efectiva que abarca el 67% de la población objetivo. Cabe destacar que en esta investigación se siguieron las directrices del Comité de Ética Científico de la universidad (http://portal.ucm.cl/comite-etica-cientifico). En este sentido, los estudiantes manifestaron voluntariamente su intención de participar en esta investigación a través de la firma de un consentimiento informado.

Técnicas e instrumentos de recopilación de datos

Los datos sobre el nivel de razonamiento estadístico de los estudiantes encuestados acerca de pruebas de hipótesis fueron recopilados a través de un instrumento adecuadamente diseñado y validado por una triangulación de 3 expertos con más de 15 años de experiencia en formación de profesores, docencia e investigación de área de estadística y didáctica de la matemática.

Este instrumento, que se muestra en la Tabla 3, consiste en un cuestionario de 8 ítems de selección simple, acompañado de una columna adyacente para la justificación de la opción elegida. Se adaptaron 7 ítems de opción simple de las investigaciones de Vallecillos (1996) y un ítem de Inzunsa y Jiménez (2013) con el objetivo de evaluar si los estudiantes chilenos tienen las mismas concepciones erróneas y por ende cometen los mismos errores mencionados en la sección anterior.

Tabla 3

Instrumento utilizado.

Ítem	Respuesta y justificación
1. ¿Cuál de las siguientes no es una hipótesis nula bien enunciada? *a) H_o: μ_x=10* b) H_o: σ_x= 3 c) H_o: $\bar{x}$= 35 d) H_o: μ_1=μ_2	
2. En una prueba de hipótesis, la potencia de la prueba es la probabilidad de: a) Rechazar la hipótesis nula, siendo ésta cierta. b) Rechazar la hipótesis nula, siendo ésta falsa. c) Que la hipótesis alternativa sea verdadera. d) Que la hipótesis nula sea falsa.	
3. Un profesor investigador siempre usa 0.05 como nivel de significación en sus estudios en inferencia estadística. Esto significa que: a) Habrá rechazado indebidamente la hipótesis nula sólo 5 de cada 100 veces b) 5 de cada 100 veces que rechace una hipótesis nula se habrá equivocado. c) Habrá aceptado una hipótesis nula falsa 95 de las 100 veces. d) 5 de cada 100 veces rechazará la hipótesis nula.	

4. Se suponen las siguientes hipótesis
H_o: $\mu_x = 50$
H_1: $\mu_x > 50$
H_2: $\mu_x < 50$ Con un nivel de significación α= 0.05, $z_{0.975}$= 1.96, un valor de $\bar{x}$= 60, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 4$, un valor del estadístico de prueba $z_c = \frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}}$ y una población normalmente distribuida, entonces se podría:
a) No rechazar H_o: $\mu_x = 50$. b) Rechazar H_o: $\mu_x = 50$.
c) Se necesita más información. d) Aceptar *H_2: $\mu_x < 50$.*
5. ¿Qué se puede concluir si el resultado de una hipótesis es estadísticamente significativo?
a) El resultado es muy interesante desde el punto de vista práctico.
b) Se está equivocado.
c) La hipótesis alternativa es probablemente correcta.
d) La hipótesis nula es probablemente correcta.

6. ¿Qué sucede cuando aumenta el nivel de significación de 0,01 a 0,05?
a) Menos probabilidad de cometer error de tipo I (α)
b) Mayor probabilidad de cometer error tipo I (α)
c) Menos probabilidad de cometer error tipo II (β)
d) b) y c)
7. Si a un nivel de significación 0.05 la hipótesis nula no es rechazada, entonces ¿qué puede decirse sobre la probabilidad de cometer error de tipo II?
a) Es igual a 0,05. b) Es igual a 0,95. c) Es menor del 5%.
d) No puede ser determinada con la información anterior.
8. *Un profesor de matemáticas ha hecho una prueba a sus estudiantes de su curso de Geometría de Pedagogía en Matemáticas y Computación. El profesor considera que sus alumnos se encuentran con todos los conocimientos disciplinares suficientes cuando no cometen más de 19 errores en dicha prueba. Para corroborar su conjetura tomó una muestra aleatoria de 10 estudiantes del curso y obtuvo los siguientes resultados de cada estudiante (en cantidad errores): 18, 22, 21, 19, 18, 17, 19, 20, 22, 20. Considerando que los datos se distribuyen normalmente y que: $\bar{x}$ = 19.6, s^2= 2.94, t_c=1.10, y $t_{0,95(9)}$ =1.8331. Además, se ha fijado anteriormente 5% del nivel de significación. ¿Qué conclusión cree usted que el profesor de matemáticas puede obtener?*
a) Se requiere más información. b) Podría ser que los estudiantes cometen más de 19 errores en la prueba, pero el tamaño de la muestra es demasiado pequeño para descubrirlo c) La evidencia muestral no fue suficiente para rechazar la hipótesis nula, en otras palabras, no es posible afirmar que los estudiantes cometen más de 19 errores. d) Debe aceptar la hipótesis alternativa.

Los conceptos de pruebas de hipótesis estadísticas evaluados fueron: formulación de las hipótesis, definición de la potencia de la prueba, proceso de las pruebas de hipótesis, la probabilidad de cometer error tipo I y tipo II, nivel de significación de la prueba de hipótesis estadística, distribución muestral del estadístico, parámetro y estadístico de prueba y criterio de decisión. En el cuestionario, los ítems fueron dispuestos y enumerados en concordancia con su nivel de complejidad, según Tabla 2 y no se consideraron ítems con nivel máximo alcanzable de abstracción extendida, esto ya que por definición no se consigue alcanzar este nivel dentro del contexto de aprendizaje universitario.

Por otro lado, siguiendo a Aravena y Caamaño (2013), diseñamos una matriz de análisis pre-armada (Tabla 4) para categorizar el nivel de razonamiento de los estudiantes en función de los descriptores de los cuatro niveles de la taxonomía SOLO, la cual también fue validada por los mismos expertos en estadística y didáctica de la matemática. Cabe recalcar que por la naturaleza de la construcción del instrumento y de la matriz de análisis, es simple e inmediato identificar el nivel de razonamiento estadístico de los estudiantes acerca de las pruebas de hipótesis mediante su respuesta y justificación en el cuestionario.

Tabla 4

Matriz de análisis de los ítems del instrumento utilizado (ítems 1-8)

Ítems de opciones (1-8)		
Ítems de nivel uni-estructural		
Aspectos relevantes	**Pre-estructural**	**Uni-estructural**
Ítem 1 Formulación de hipótesis	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	La justificación es realizada con un lenguaje estadístico adecuado y la hipótesis nula es definida en función de un parámetro. Esto muestra que el estudiante es capaz de determinar la opción correcta y descartar los distractores con fundamento.
Ítem 2 Definición de la potencia de las pruebas de hipótesis.	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	La justificación es realizada con un lenguaje estadístico adecuado y basada en la definición de la potencia de una prueba. Esto muestra que el estudiante es capaz de determinar la opción correcta y descartar los distractores con fundamento.
Ítem 3 Nivel de significación de la prueba de hipótesis.	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	La justificación es realizada con un lenguaje estadístico adecuado y es basada en el nivel de significación (tratado como la probabilidad de rechazar la hipótesis nula siendo ésta verdadera) e interpretarlo adecuadamente en contexto de la hipótesis nula.

Ítems de nivel multi-estructural

Aspectos relevantes	**Pre-estructural**	**Uni-estructural**	**Muti-estructural**
Ítem 4 **1)** Nivel de significación y criterio de decisión. **2)** Nivel de significación y distribución muestral del estadístico.	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	Se observa un manejo aislado en la obtención del estadístico de prueba observado y la toma de una decisión adecuada con respecto a la hipótesis nula.	La justificación es realizada con un lenguaje estadístico adecuado y es basada en un cálculo correcto del valor de estadístico de prueba observado y discernimiento claro entre la región de rechazo y de aceptación que permite tomar una decisión adecuada con respecto a la hipótesis nula.
Ítem 5 **1**) Definición de la hipótesis nula y alternativa **2)** Nivel de significación y distribución muestral del estadístico.	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	Presenta una confusión sobre el rechazo de la hipótesis nula.	La justificación es realizada con un lenguaje estadístico adecuado y ratifica que un resultado estadísticamente significativo es aquél que rechaza la hipótesis nula pero no comprueba en forma absoluta la hipótesis alternativa.
Ítem 6 **1)** Probabilidad de cometer error tipo I y tipo II **2)** La relación entre ellas y la potencia de la prueba.	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	Presenta una confusión sobre la relación entre el error de tipo I y II.	La justificación es realizada con un lenguaje estadístico adecuado y se basa en una interpretación correcta del nivel de significación y la relación entre el error de tipo I y II.

Ítems de nivel relacional				
Aspectos relevantes	**Pre-estructural**	**Uni-estructural**	**Muti-estructural**	**Relacional**
Ítem 7 **1)** Probabilidad de cometer error tipo I y II. **2)** La relación entre ellas. **3.)** Parámetro y estadístico de prueba.	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	Entrega la definición de la probabilidad de error de tipo II correcta, pero no puede interpretarla en función del nivel de significación.	La justificación está hecha utilizando un lenguaje estadístico adecuado y utiliza la definición correcta de la probabilidad de error de tipo II pero tiene error o confusión a la hora de interpretarla en función del nivel de significación.	La justificación es hecha utilizando un lenguaje estadístico adecuado y logra interpretar correctamente la definición de la probabilidad de error tipo II en función del nivel de significación.

Ítem 8 **1)** Nivel de significación y criterio de decisión. **2)** Distribución del estadístico de prueba. **3)** Tamaño de la muestra. **4)** Parámetro y estadístico de prueba	No responde o presenta una justificación incompleta o no concordante con lo que pide el ítem.	Presenta la formulación correcta de la hipótesis o cálculo correcto del valor estadístico de prueba.	La justificación se basa en la formulación correcta de la hipótesis, cálculo correcto del valor del estadístico de prueba, pero comete error en la toma de decisión por una confusión entre las regiones de rechazo y aceptación.	La justificación se basa en la formulación correcta de la hipótesis, en el cálculo correcto del valor del estadístico de prueba y en el empleo adecuado de la región de rechazo y aceptación para tomar una decisión sobre la hipótesis nula.

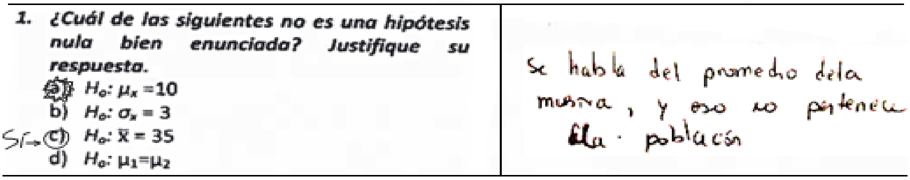
ANÁLISIS Y RESULTADOS

En esta sección, se analiza el nivel de razonamiento de los estudiantes, utilizando la matriz de análisis presentada en Tabla 4. De acuerdo a los resultados obtenidos en cada ítem, se obtiene una conclusión general sobre el nivel de razonamiento estadístico predominante en los individuos en estudio.

En el ítem 1 se evalúa la formulación de las hipótesis estadísticas con el propósito de conocer si los estudiantes establecen dicha hipótesis en función de un parámetro poblacional. Para evidenciar el nivel de razonamiento estadístico en este ítem se ha elegido un ejemplo de las respuestas y justificaciones típicas entregados por los alumnos, como se muestra en Figura 1. En esta figura se puede visualizar que el estudiante distinguió correctamente el estadístico del parámetro poblacional y reconoció la notación habitual. En este ítem, el 72,41% de los estudiantes respondieron y justificaron correctamente como en Figura 1, mientras que el 27,59% de los estudiantes evidencian confusión para definir la hipótesis en función de un parámetro poblacional y no de un estadístico muestral.

Figura 1

Respuesta y justificación típica del ítem1.



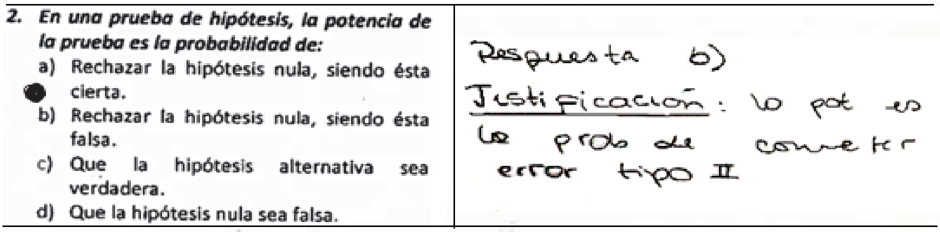
1. ¿Cuál de las siguientes no es una hipótesis nula bien enunciada? Justifique su respuesta.
a) H_o: $\mu_x = 10$
b) H_o: $\sigma_x = 3$
Sí→ c) H_o: $\bar{X} = 35$
d) H_o: $\mu_1 = \mu_2$

Se habla del promedio dela muestra, y eso no pertenece a la población

En el ítem 2, el contenido evaluado fue la definición de la potencia de las pruebas de hipótesis. En la Figura 2 se entrega un ejemplo de las respuestas y justificaciones típicas de este ítem. En esta figura se puede visualizar que el estudiante eligió la alternativa correcta, pero con su justificación evidenció confusión entre los conceptos de potencia y probabilidad de cometer error del tipo II. Cabe destacar que en este ítem, el 79,31% de todas las respuestas fueron clasificadas en el nivel pre-estructural.

Figura 2

Respuesta y justificación no adecuada del ítem 2.



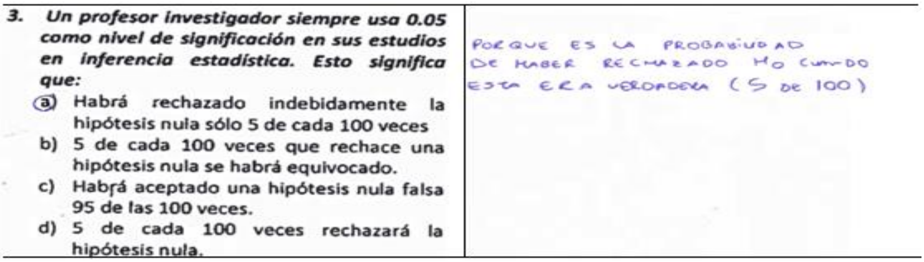
2. En una prueba de hipótesis, la potencia de la prueba es la probabilidad de:
a) Rechazar la hipótesis nula, siendo ésta cierta.
b) Rechazar la hipótesis nula, siendo ésta falsa.
c) Que la hipótesis alternativa sea verdadera.
d) Que la hipótesis nula sea falsa.

Respuesta b)
Justificación: la pot es la prob de cometer error tipo II

En el ítem 3, se evaluó el nivel de significación de las pruebas de hipótesis con el propósito de analizar la comprensión de este concepto. Figura 3 muestra un ejemplo de una respuesta y justificación correcta del ítem, donde se puede apreciar que el estudiante manejó correctamente el concepto del nivel de significación. En este ítem el 79,31% de las respuestas de los estudiantes evidenciaron confusión en el concepto de nivel de significación y sólo 20,69% evidenciaron el manejo correcto de este concepto como en Figura 3.

Figura 3

Respuesta y justificación correcta del ítem 3.

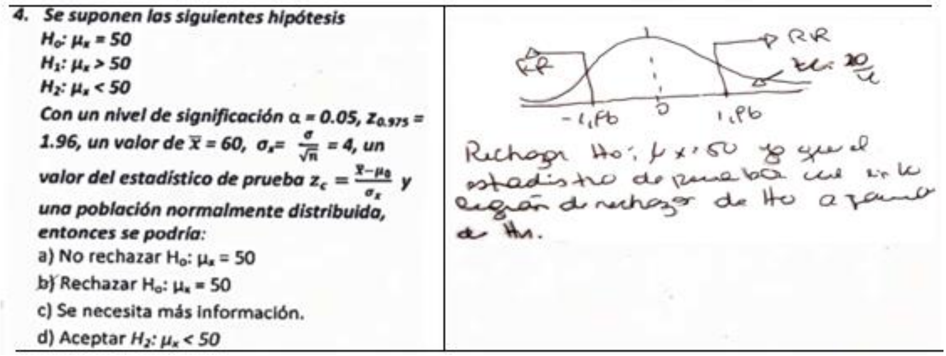


3. *Un profesor investigador siempre usa 0.05 como nivel de significación en sus estudios en inferencia estadística. Esto significa que:* (a) Habrá rechazado indebidamente la hipótesis nula sólo 5 de cada 100 veces b) 5 de cada 100 veces que rechace una hipótesis nula se habrá equivocado. c) Habrá aceptado una hipótesis nula falsa 95 de las 100 veces. d) 5 de cada 100 veces rechazará la hipótesis nula.	PORQUE ES LA PROBABILIDAD DE HABER RECHAZADO H_O CUANDO ESTA ERA VERDADERA (5 DE 100)

En el ítem 4 se evaluó el nivel de significación y criterio de decisión, y el nivel de significación y distribución muestral del estadístico. En la Figura 4 se entrega un ejemplo de una respuesta y justificación correcta, que corresponden al 31,83% de todas las respuestas del ítem.

Figura 4

Respuesta y justificación correcta del ítem 4.



4. *Se suponen las siguientes hipótesis* $H_0: \mu_x = 50$ $H_1: \mu_x > 50$ $H_2: \mu_x < 50$ ***Con un nivel de significación $\alpha = 0.05$, $z_{0.975} = 1.96$, un valor de $\overline{x} = 60$, $\sigma_x = \frac{\sigma}{\sqrt{n}} = 4$, un valor del estadístico de prueba $z_c = \frac{\overline{x}-\mu_0}{\sigma_x}$ y una población normalmente distribuida, entonces se podría:*** a) No rechazar $H_0: \mu_x = 50$ b) Rechazar $H_0: \mu_x = 50$ c) Se necesita más información. d) Aceptar $H_2: \mu_x < 50$	Rechazo Ho: μx = 50 ya que el estadistico de prueba cae en la region de rechazo de Ho a favor de H1.

En el ítem 5, se evaluó la definición de las hipótesis nula y alternativa, y el nivel de significación. Para mostrar el nivel de razonamiento estadístico de los estudiantes en este ítem, en la Figura 5 se entrega un ejemplo de las respuestas típicas de los estudiantes. En esta figura, se puede observar que se entregó una respuesta incorrecta y una justificación no concordante con lo que se pidió en el ítem. Cabe destacar que, en este ítem, el 86% de las respuestas entregadas son incorrectas, clasificándolos en nivel pre-estructural.

Figura 5

Respuesta y justificación no concordante del ítem 5.

5. ¿Qué se puede concluir si el resultado de una hipótesis es estadísticamente significativo? a) El resultado es muy interesante desde el punto de vista práctico b) Se está equivocado c) La hipótesis alternativa es probablemente correcta d) La hipótesis nula es probablemente correcta	La hipótesis nula es el objeto de estudio

En el ítem 6, se evaluaron los conceptos de probabilidad de cometer error de tipo I y II, y la relación entre ellas. Figura 6 muestra un ejemplo de las respuestas típicas entregadas en este ítem, donde se puede visualizar que el estudiante, a pesar de realizar una respuesta acertada, dejó el ejercicio sin justificar mencionando sólo la definición de error tipo I y II. En este ítem, la tendencia es recordar la definición de error tipo I y/o II pero no la relación entre ellos, por ende el 44,83% de todas las respuestas entregadas fueron clasificadas en nivel uni-estructural.

Figura 6

Respuesta acertada y justificación incompleta del ítem 6.

6. ¿Qué sucede cuando aumenta el nivel de significación de 0.01 a 0.05? a) Menos probabilidad de cometer error de tipo I (α) b) Mayor probabilidad de cometer error tipo I (α) c) Menos probabilidad de cometer error tipo II (β) d) b) y c)	Tipo I: Rechazar Ho cuando es V Tipo II No rechazar Ho cuando es F.

En el ítem 7, se evaluaron contenidos como probabilidad de cometer error de tipo I y II, la relación entre ellas, y el concepto de parámetro y estadístico de prueba. En la Figura 7 se entrega una de las respuestas y justificaciones típicas donde se puede observar que el estudiante no sólo seleccionó mal la alternativa, sino que entregó una justificación equivocada. Cabe destacar que el 72,41% de las respuestas siguen esta tendencia, evidenciando bastante confusión en los contenidos por lo que fueron clasificados en un nivel pre-estructural.

Figura 7

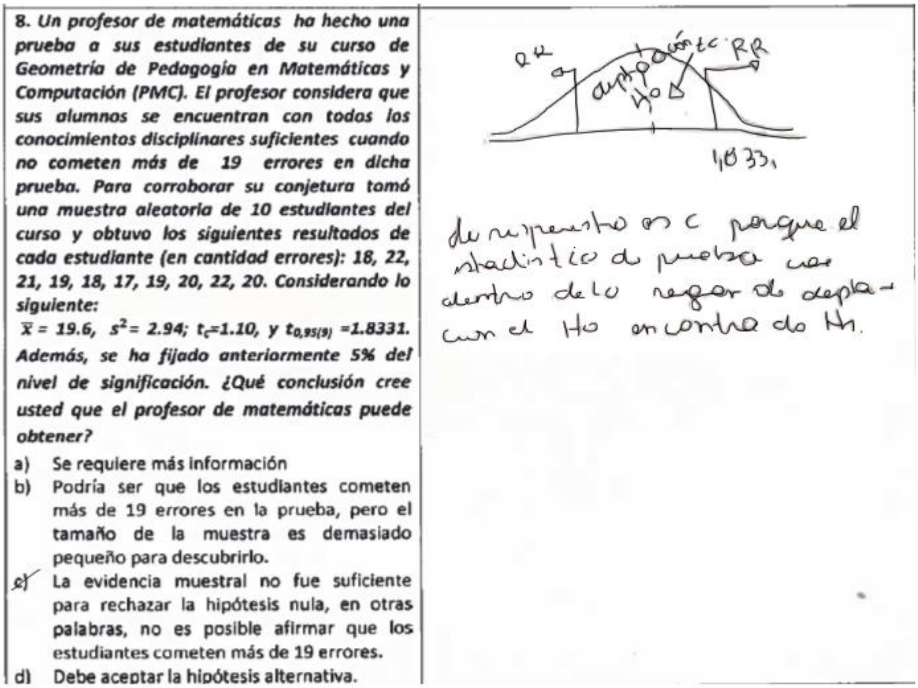
Respuesta y justificación no concordante del ítem 7.

7. Si a un nivel de significación 0.05 la hipótesis nula no es rechazada, entonces ¿qué puede decirse sobre la probabilidad de cometer error de tipo II? a) Es igual a 0.05 b) Es igual a 0.95 c) Es menor del 5% d) No puede ser determinada con la información anterior.	El error del tipo II es el complemento del error del tipo 1.

En el ítem 8, los contenidos evaluados fueron nivel de significación y criterio de decisión, distribución del estadístico de prueba, tamaño de la muestra, y parámetro y estadístico de prueba. En este ítem, el 58,62% de las respuestas corresponden a un nivel pre-estructural y 20,69% a uno relacional. Para evidenciar esta situación, en la Figura 8 se muestra una respuesta y justificación correcta del ítem, donde se puede visualizar que el estudiante llevó los datos entregados en el enunciado a un gráfico obteniendo visualmente el valor del estadístico de prueba en la región de rechazo, por lo tanto, concluyó que no hubo evidencia suficiente para rechazar H_0.

Figura 8

Respuesta y justificación correcta del ítem 8.


8. Un profesor de matemáticas ha hecho una prueba a sus estudiantes de su curso de Geometría de Pedagogía en Matemáticas y Computación (PMC). El profesor considera que sus alumnos se encuentran con todos los conocimientos disciplinares suficientes cuando no cometen más de 19 errores en dicha prueba. Para corroborar su conjetura tomó una muestra aleatoria de 10 estudiantes del curso y obtuvo los siguientes resultados de cada estudiante (en cantidad errores): 18, 22, 21, 19, 18, 17, 19, 20, 22, 20. Considerando lo siguiente:

$\bar{x} = 19.6$, $s^2 = 2.94$; $t_c=1.10$, y $t_{0,95(9)} = 1.8331$.

Además, se ha fijado anteriormente 5% del nivel de significación. ¿Qué conclusión cree usted que el profesor de matemáticas puede obtener?

a) Se requiere más información
b) Podría ser que los estudiantes cometen más de 19 errores en la prueba, pero el tamaño de la muestra es demasiado pequeño para descubrirlo.
c) La evidencia muestral no fue suficiente para rechazar la hipótesis nula, en otras palabras, no es posible afirmar que los estudiantes cometen más de 19 errores.
d) Debe aceptar la hipótesis alternativa.

A modo de resumen, en la Tabla 5 se entrega el porcentaje de estudiantes clasificados en cada uno de los niveles de razonamiento estadístico según la taxonomía SOLO.

Tabla 5

Porcentaje de los niveles de razonamiento de taxonomía SOLO de los estudiantes en cada ítem.

Ítem	Pre-estructural	Uni-estructural	Multi-estructural	Relacional	Nivel predominante
1	27,59	72,41	-----	-----	uni-estructural
2	79,31	20,69	-----	-----	pre-estructural
3	72,41	27,59	-----	-----	pre-estructural
4	44,83	24,14	31,03	-----	pre-estructural

5	86	7	7	-----	pre-estructural
6	41,38	44,83	13,79	-----	uni-estructural
7	72,41	20,69	0	6,9	pre-estructural
8	58,62	13,79	6,9	20,69	pre-estructural

En base a los resultados de los 8 ítems del instrumento y Tabla 5, se puede afirmar que en las respuestas de los estudiantes de Pedagogía en Matemática acerca de las pruebas de hipótesis estadísticas predominaron las no concordantes o incompletas, por lo tanto, basados en estos datos, se puede concluir que los estudiantes de Pedagogía en Matemática tuvieron el nivel de razonamiento estadístico pre-estructural y uni-estructural de la taxonomía SOLO acerca de las pruebas de hipótesis estadísticas.

CONCLUSIONES

De acuerdo con lo comentado en la sección Análisis y Resultados, se evidenció que los estudiantes tuvieron dificultad en los conceptos esenciales de las pruebas de hipótesis, sobre todo en relacionar los conceptos para obtener una conclusión en contexto del problema planteado que permite tomar una decisión correcta. Casi en todas las respuestas del cuestionario los estudiantes mostraron un manejo aislado y superficial de los conceptos de pruebas de hipótesis e incluso seleccionando la alternativa correcta no pudieron justificarlo con un adecuado lenguaje estadístico. También se observaron dificultades para tomar una decisión correcta y relatar una conclusión contextualizada adecuada, resultados similares a los obtenidos por Castro et al. (2007) y Batanero (2013a).

En cuanto al manejo conceptual que poseen los estudiantes de Pedagogía en Matemática sobre este tema, hubo un alto porcentaje de estudiantes que establecieron la hipótesis en base a los parámetros poblaciones, sin embargo, se registró una minoría que las formuló en función de un estadístico muestral como lo observado en las investigaciones de Vallecillos (1996), Vallecillos y Batanero (1997) e Inzunsa y Jiménez (2013). Además, se corroboró al igual que en los estudios de Batanero et al. (1994), Vallecillos (1996), Vallecillos y Batanero (1997), Castro et al. (2007), Batanero et al. (2012) e Inzunsa y Jiménez (2013) que una gran parte de estudiantes incurrieron en error al interpretar el nivel de significación como error tipo I y no probabilidad de éste. También se observó una confusión con la definición del nivel de significación, dificultad para determinar la región crítica y de aceptación, impidiéndoles tomar una decisión adecuada en el ítem 4 como lo obtenido en Vallecillos y Batanero (1997).

Por otro lado, en el ítem 5 y 8, los estudiantes no distinguieron entre la hipótesis nula y alternativa ni asimilaron que el objetivo de la prueba es rechazar la hipótesis nula cuando el resultado es estadísticamente significativo. En el ítem 6, aunque su objetivo era estudiar la relación entre los errores de tipo I y II al aumentar el nivel de significación, en la columna de justificación se observaron que muchos estudiantes interpretaron los errores tipo I y II como probabilidad de sucesos complementarios como señalan Vallecillos (1996) y Batanero (2013b).

Entonces, en base a los resultados obtenidos realizamos las siguientes conclusiones y exponemos algunas limitaciones:

1) Los estudiantes de Pedagogía en Matemática de una Universidad chilena se encuentran en el nivel de razonamiento estadístico pre-estructural y uni-estructural de la taxonomía SOLO acerca de las pruebas de hipótesis estadísticas, confirmando la hipótesis formulada al inicio de esta investigación.

2) Las pruebas de hipótesis son un tema conceptual y procedimentalmente difícil para los estudiantes de Pedagogía en Matemática que aprobaron los cursos de estadística I y II.

3) Los resultados obtenidos evidencian que los estudiantes no lograron los resultados de aprendizaje esperados del programa de estudio, esto sirve como insumo para sugerir a la carrera la necesidad de enmendar esta insuficiencia mediante herramientas remediales antes de que estos estudiantes egresen y se inserten en la labor docente.

4) Por último, aunque no es posible generalizar los resultados obtenidos para los estudiantes de pedagogía en Matemática en formación de todas las Universidades de Chile, se podría conjeturar que, al aplicar el instrumento en otras universidades chilenas, los estudiantes obtendrían resultados similares a los obtenidos en la presente investigación.

CONTRIBUCIONES ESTADOS DE LOS AUTORES

C.S concibió la idea presentada. C.S y C.M llevaron a cabo la investigación y la recolección de datos. Ambas autoras analizaron los datos y participaron activamente en la discusión de los resultados, revisando y obteniendo la versión final del manuscrito.

DECLARACIÓN DE DISPONIBILIDAD DE DATOS

Los datos que respaldan los resultados del presente estudio estarán disponibles por los autores, previa solicitud razonable.

REFERENCIAS

Amaro, J. y Sánchez, E. (2015). La toma de decisiones en una situación de riesgo. In: *XIV Conferencia Interamericana de Educación Matemática*.

Aravena, M., y Caamaño, C. (2013). Niveles de razonamiento geométrico en estudiantes de establecimientos municipalizados de la Región del Maule: Talca, Chile. *Revista latinoamericana de investigación en matemática educativa*, *16*(2), 179-211.

Batanero, C. (2000) ¿Hacia dónde va la educación estadística? *Blaix*, *15*(2), 2-13.

Batanero, C. (2005) Statistics education as a field for research and practice. In: *Proceedings of the 10th international commission for mathematical instruction*. Copenhagen, Denmark: International Commission for Mathematical Instruction.

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

Batanero, C. (2013b) Sentido estadístico. Componentes y desarrollo. En: *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria*, *1* (p. 55-61). Granada: Universidad de Granada.

Batanero, C., Godino, J.,Vallecillos, A, Green, D. y Holmes, P. (1994) Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, *25*(4), 527-547.

Batanero, C., Vera, O. y Díaz, C. (2012) Dificultades de estudiantes de Psicología em la comprensión del contraste de hipótesis. *Números. Revista de Didáctica de las Matemáticas*, *80*, 91-101.

Biggs, J.B. y Collis, K.F. (1982) *Evaluating the quality of learning. The SOLO taxonomy* (Structure of the Observed Learning Outcome). Academic Press.

Castro, S., Vanhoof, S., Van Den Noortgate y W. Onghena, P. (2007) Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98-113.

Del Pino, G. y Estrella, S. (2012) Educación estadística Relaciones con la matemática. *Pensamiento Educativo: Revista de Investigación Educacional Latinoamericana*, *49*(1), 53-64.

Estrella, S. (2008) Medidas de tendencia central en la enseñanza básica en Chile: análisis de un texto de séptimo año. *Revista Chilena de Educación Matemática (RECHIEM)*, *4*(1), 20-32.

Estrella, S. (2014) Un imperativo moral: la enseñanza de la estadística no puede dejarse al azar. In: *Encuentro colombiano de Educación Estocástica. 1.* Memorias (pp. 67-77) Bogotá: Asociación Colombiana de Educación Estocástica.

Gal, I. (2002) Adults' statistical literacy: Meanings, components, responsibilities. *International statistical review*, *70*(1), 1-25.

García, J. y Sánchez, E. (2013) Niveles de razonamiento probabilístico de estudiantes de bachillerato frente a una situación básica de variable aleatoria y distribución. *Probabilidad Condicionada: Revista de didáctica de la Estadística*, *2*, 417-424.

García, J. y Hernández, E. (2018) Niveles de razonamiento probabilístico de estudiantes de bachillerato sobre la noción de la distribución binomial. *Acta Latinoamericana de Matemática Educativa, 31*, 962-969.

Garfield, J. (2002) The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, *10*(3). 1-12

Hacking, I. (1990) *The taming of chance*. Cambridge, MA: Cambridge University Press.

Huerta, P. **(**1997) *Los niveles de van Hiele en relación con la taxonomía SOLO y en los mapas conceptuales.* Tesis (Doctorado en Matemática) – Departamento de Didáctica de la Matemática, Universidad de Valencia, Valencia.

Inzunsa, S. y Jiménez, J. (2013) Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. *Revista latinoamericana de investigación en matemática educativa*, *16*(2), 179-211.

Leenen, I. (2012) La prueba de la hipótesis nula y sus alternativas: revisión de algunas críticas y su relevancia para las ciencias médicas. *Investigación en educación médica*, 4, 225-234.

MINEDUC (2009a) *Propuesta Ajuste Curricular: Objetivos Fundamentales y Contenidos Mínimos Obligatorios*. Ministerio de Educación de Chile.

MINEDUC (2009b) *Fundamentos del Ajuste Curricular en el Sector de Matemática.* Ministerio de Educación de Chile.

MINEDUC (2009c). *Mapas de progreso del aprendizaje*. Sectormatemática. Mapa de progreso de datos y azar. Ministerio de Educación de Chile.

Montgomery, D. y Runger, G. (2012) *Probabilidad y Estadística Aplicadas a la Ingeniería*. Limusa and Wiley.

Muñoz, C. (2015) Caracterización del razonamiento estadístico sobre el concepto de estimación puntual en estudiantes de grado noveno. In: P. Scott y A. Ruiz (Eds.). *Estadística y Probabilidad* (pp. 20-32*)* Comité Interamericano de Educación Matemática.

Pfannkuch, M. (2005a) Characterizing year 11 student´s evaluation of a statistical process. *Statistics Education Research Journal, 4*(2), 5-25. Pfannkuch, M. (2005b) Probability and statistical inference: how can teachers enable learners to make the connection? In *Exploring probability in school* (pp. 267-294). Springer.

Reading, Ch. y Reid, J. (2006) An emerging hierarchy of reasoning about distribution: from a variation perspective. *Statistics Education Research Journal*, *5*(2), 46-68.

Vallecillos, A. y Batanero, C. (1997) Aprendizaje y enseñanza del contraste de hipótesis: concepciones y errores. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, *15*(2), 189-197.

Vallecillos, A. (1996) *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadístico*. Comares.

Vásquez, C. y Alsina, Á. (2017) Lenguaje probabilístico: un camino para el desarrollo de la alfabetización probabilística. Un estudio de caso en el aula de Educación Primaria. *Boletim de Educação Matemática*, *31*(57), 454-478.

Watson, J. (1997) Assessing statistical thinking using the media. In I. Gal and J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). IOS.