

Diagnóstico sobre el razonamiento inferencial en estudiantes de licenciatura del área económico-administrativa

Diagnosis about inferential reasoning in undergraduate students of the economic management area

Salvador Sandoval Bravo
Pedro Luis Celso-Arellano
Víctor Hugo Gualajara Estrada

RESUMEN

Entre las líneas de investigación vanguardista en el ámbito de la educación estadística se encuentra el razonamiento inferencial, como herramienta fundamental para la toma de decisiones; factor que acentúa su importancia tratándose de licenciaturas del área económico-administrativa donde los futuros profesionales deben probar hipótesis, construir intervalos de confianza y hacer estimaciones sobre poblaciones basándose en datos referentes a muestras representativas. En este sentido, el objetivo de este trabajo es medir el razonamiento inferencial de los estudiantes de licenciatura del área económico-administrativa de la Universidad de Guadalajara para diagnosticar e identificar fortalezas y áreas de oportunidad, que puedan utilizarse en la implementación de soluciones pertinentes tanto en el aspecto pedagógico como disciplinar, y así elevar el rendimiento futuro de los alumnos. Para ello se aplicó el examen CAOS-4 a una muestra de 326 estudiantes, dicho instrumento ha sido validado estadísticamente para estudiantes universitarios, y permite determinar el nivel de razonamiento inferencial. Los resultados encontrados muestran que los estudiantes tienen dificultades para identificar y plantear hipótesis, no entienden la naturaleza probabilista de las conclusiones de la inferencia estadística, interpretan erróneamente los intervalos de confianza y el margen de error. De esta manera, la enseñanza de la estadística a nivel licenciatura debe enfatizar y reforzar el razonamiento inferencial, puesto que de 21 reactivos de esta índole medidos, solo tres de ellos mostraron un porcentaje aprobatorio de respuestas correctas, mientras que las demás mostraron un porcentaje deficiente o muy deficiente.

Palabras clave: razonamiento estadístico, educación universitaria, evaluación del aprendizaje.

ABSTRACT

Within the forefront research areas in the field of statistical education is located the inferential reasoning, as a fundamental tool for decision-making; factor that accentuates its importance in the case of economics and management careers where future professionals must test hypotheses, build confidence intervals and make estimates on populations based on data referring to representative samples. In this sense, the aim of this work is to measure the inferential reasoning of undergraduate students of the Economics-Management area of the University of Guadalajara to diagnose and identify strengths and opportunity areas, which can be used in the implementation of pertinent solutions both in the pedagogical and disciplinary aspects, and thus increase the future performance of students. For this, the CAOS-4 exam was applied to a sample of 326 students; such instrument has been statistically validated for university students and allows determining the level of inferential reasoning. The results found show that students have problems in identifying and formulating hypotheses, they do not understand the probabilistic nature of the conclusions of statistical inference, they misinterpret the confidence intervals and the margin of error. In this way, the teaching of statistics at the undergraduate level should emphasize and reinforce inferential reasoning, since, from 21 items of this kind measured, only 3 of them showed a passing percentage of correct answers, while the others showed a poor or very poor percentage.

Keywords: statistical reasoning, undergraduate education, learning assessment.

INTRODUCCIÓN

La educación estadística en general y el razonamiento inferencial en particular juegan un papel importante en el mundo de los negocios y la economía, donde la generación incesante de información nueva marca el ritmo de los cambios y el progreso, ya que a través de ella se pueden inferir conclusiones sobre una población con datos de una muestra e interpretarlos de manera adecuada. En este sentido, en México los estudiantes universitarios de las carreras del área económico-administrativa llevan por lo menos un curso de estadística, el cual puede incluir estadística descriptiva e inferencial, o estas áreas de la estadística en asignaturas diferentes. Dicho curso debería motivar al estudiante sobre el potencial de la estadística y prepararlo para la utilización de sus métodos y procedimientos en la resolución de problemas en el mundo real (Rumsey, 2002). De hecho, la estadística es una de las materias cuantitativas más relevantes en un plan de estudios universitario (Watson, 1997), y en particular la metodología de la estadística inferencial es de capital importancia en el diseño y contrastación de experimentos, y puede ser ampliamente utilizada en muchos campos, tanto de las ciencias naturales como de las ciencias sociales (Huang, 2018); no obstante, existen consideraciones y limitaciones que deben tomarse en cuenta, sobre todo por el carácter probabilista y el margen de error asociado a sus conclusiones e interpretación de sus resultados (Hubbard, Haig y Parsa, 2019; Tong, 2019).

Salvador Sandoval Bravo. Profesor-investigador del Departamento de Métodos Cuantitativos del Centro de Ciencias Económico Administrativas de la Universidad de Guadalajara, México. Es doctor en Negocios y Estudios Económicos, maestro en Administración de Negocios y licenciado en Matemáticas. Reconocimiento Perfil PRODEP y Miembro del Sistema Nacional de Investigadores. Ha publicado libros, capítulos de libros y artículos en revistas indizadas nacionales e internacionales. Entre sus publicaciones recientes se encuentran “Análisis descriptivo de las diferencias por sexo en los estudiantes de licenciatura de la Universidad de Guadalajara” y “Production of comics in Powtoon as a teaching-learning strategy in an operations”. Correo electrónico: salvador.sandoval@academicos.udg.mx. ID: <https://orcid.org/0000-0001-9434-6536>.

Pedro Luis Celso-Arellano. Profesor-investigador del Departamento de Métodos Cuantitativos del Centro de Ciencias Económico Administrativas de la Universidad de Guadalajara, México. Es doctor en Negocios y Estudios Económicos, maestro en Negocios y Estudios Económicos y licenciado en Administración en Empresas Agropecuarias. Reconocimiento Perfil PRODEP y miembro del Sistema Nacional de Investigadores. Entre sus publicaciones recientes se encuentran “Análisis descriptivo de las diferencias por sexo en los estudiantes de licenciatura de la Universidad de Guadalajara” y “Diagnóstico sobre la lectura e interpretación de gráficos estadísticos en estudiantes de licenciatura de ciencias económica-administrativas en la Universidad de Guadalajara”. Correo electrónico: pedro.celso@academicos.udg.mx. ID: <https://orcid.org/0000-0003-3856-6344>.

Víctor Hugo Gualajara Estrada. Profesor e investigador del Departamento de Métodos Cuantitativos del Centro de Ciencias Económico Administrativas de la Universidad de Guadalajara, México. Es doctor en Ciencias Físico Matemáticas, maestro en Tecnologías de la Información y licenciado en Matemáticas. Reconocimiento al Perfil PRODEP. Ha publicado capítulos de libros y artículos en revistas indizadas nacionales e internacionales. Entre sus publicaciones recientes se encuentran “Criptomonedas en América Latina y China: el rol del e-commerce y la especulación financiera” y “Diagnóstico sobre la lectura e interpretación de gráficos estadísticos en estudiantes de licenciatura de ciencias económica-administrativas en la Universidad de Guadalajara”. Correo electrónico: victor.gualajara@academicos.udg.mx. ID: <https://orcid.org/0000-0002-5405-3779>.

En este sentido, a partir de información recolectada sobre una muestra, los estudiantes universitarios emplean la estadística inferencial para analizar e interpretar dichos datos y así responder a preguntas planteadas o a contrastar hipótesis referentes a una población específica, en otras palabras, a realizar inferencias estadísticas en el sentido literal de la disciplina. Así que la estadística inferencial es un instrumento cuantitativo indispensable que apoya al estudiante a resolver problemas, y partir de ella dar soluciones en este mundo de la era de la información y la tecnología (Noll, Gebresenbet y Glover, 2016; Makar y Rubin, 2017; Peñaloza y Vargas, 2017).

Por otro lado, el aprendizaje de la estadística inferencial en las carreras del ámbito económico-administrativo privilegia el aspecto práctico de la materia sobre el aspecto teórico y formal. De esta manera los alumnos, además de estudiar la parte conceptual y analítica de la materia, aprenden, sobre todo, a resolver problemas de utilidad práctica relacionados con el mundo de los negocios y la economía, lo cual logran aplicando los métodos y procedimientos de la estadística inferencial. Sin embargo, cuando los egresados de estas carreras se enfrentan a situaciones y problemas reales en el ejercicio profesional, es frecuente que se confundan al identificar la prueba o técnica estadística más adecuada en el contexto en que se estén desempeñando (Beitz, 1998; Zellner, Boerst y Tabb, 2007). De acuerdo a Gibbs, Shafer y Miles (2017); la confusión del alumno puede deberse, entre otras razones, a la forma deficiente en que aprendió la estadística inferencial durante su vida académica.

Tradicionalmente, el aprendizaje en el aula se puede lograr a través de la resolución de problemas estadísticos y estudios de casos (ya sea de manera puramente analítica o utilizando *software* especializado), aplicación de exámenes, entre otros (Delucchi, 2014). En este sentido, Reaburn (2018) afirma que es frecuente que los estudiantes aprueben los cursos de estadística inferencial sin tener la capacidad de explicar los principios detrás de esta disciplina; en otras palabras, los estudiantes solo aplicaban conocimiento procedimental (aprendizaje superficial y mecánico) sin el conocimiento conceptual asociado.

Por otro lado, Doerr, DelMas y Makar (2017) enfatizan la importancia del razonamiento sobre modelos en la inferencia estadística, y discuten enfoques pedagógicos para construir puentes entre los datos y la perspectiva probabilística en el contexto de la inferencia estadística. En este sentido, Pfannkuch y Wild (2015) describen las razones por las cuales la inferencia estadística y el razonamiento inferencial deben ser parte de las experiencias de aprendizaje de los estudiantes, también discuten sobre los fundamentos conceptuales necesarios para una introducción por etapas a la inferencia estadística; la implementación de tal vía en las aulas requirió el desarrollo de nuevas visualizaciones dinámicas, verbalizaciones, formas de razonamiento, trayectorias de aprendizaje, diseños de nuevos recursos y material didáctico, entre otros.

En este tenor, Weinberg, Wiesner y Pfaff (2010) afirman que el razonamiento inferencial es un componente central de la estadística, y sugieren que los estudiantes deben desarrollar una comprensión informal de las ideas que subyacen a la inferencia al mismo tiempo que la formalización conceptual; así, cuando resuelven problemas prácticos visualizan de manera general su contexto, tanto el aspecto teórico como el aspecto procedimental. Por lo tanto, el razonamiento inferencial favorece la comprensión integral de los conceptos y la metodología de la inferencia estadística, y no solo la aplicación automática de procedimientos y algoritmos para la resolución de problemas.

Por otro lado, a nivel mundial, en los últimos 50 años se ha promovido la educación estadística en general, y el razonamiento inferencial en particular, en todos los niveles educativos desde el nivel elemental hasta el nivel superior, a través de programas tales como Schools Council Project on Statistical Education en Inglaterra, o Quantitative Literacy, Data Driven Curriculum Strand for High School Mathematics y National Council of Teachers of Mathematics en Estados Unidos. Estos proyectos han incorporado de manera constante propuestas vanguardistas en el ámbito del razonamiento estadístico y el razonamiento inferencial (Cuevas e Ibáñez, 2008).

En este sentido, en años recientes se estableció el International Statistical Literacy Project, un plan a escala global vigente, cuyos objetivos incluyen la generación de planteamientos de investigación en el área de estadística educativa en general, entre las cuales se encuentran propuestas vanguardistas en el ámbito del razonamiento inferencial (Sánchez, 2010). Por otro lado, existen programas a nivel nacional que se desarrollan en diversos países, tales como Chile, Argentina, China, Australia y Nueva Zelanda, naciones que tradicionalmente han promovido la educación estadística en todos sus niveles educativos y que han integrado los últimos avances que las investigaciones han reportado a sus planes de estudio, incluyendo, claro está, las que se refieren al razonamiento inferencial.

En México existe la Asociación Mexicana de Estadística (AME). La AME tiene como objetivos principales: a) fortalecer el ámbito académico de la divulgación y análisis de los planteamientos académicos y conclusiones de investigación en materia de educación estadística; b) coordinar profesores, investigadores y cuerpos colegiados nacionales para identificar las principales áreas de oportunidad de la investigación y educación estadística en el país, y c) fomentar el trabajo conjunto con instituciones académicas internacionales en el establecimiento de una agenda común que promueva nuevos desarrollos e investigaciones en la educación estadística (AME, 2020).

A nivel superior, debido al carácter autónomo de las universidades, son estas las que definen los contenidos temáticos que conformarán la asignatura o asignaturas de la disciplina estadística (Cuevas e Ibáñez, 2008). En este sentido, la responsabilidad de evaluar el aprendizaje estadístico de los estudiantes universitarios también es facultad de las universidades.

De esta manera, en el Centro Universitario de Ciencias Económico Administrativas (CUCEA) de la Universidad de Guadalajara (U de G), que imparte programas de licenciatura y posgrado del área de economía y negocios, la aplicación de exámenes departamentales para medir el razonamiento inferencial de los alumnos de licenciatura en la materia de estadística inferencial está a cargo de la Academia de Estadística.¹ A este respecto, los resultados de los exámenes departamentales para la materia de estadística inferencial (en una escala de 0 a 100), en el segundo semestre del 2017, primer semestre del 2018, segundo semestre del 2018 y primer semestre del 2019 fueron 45.21, 47.31, 45.86 y 48.42 respectivamente. Puede observarse que dichos promedios son muy deficientes, y han oscilado alrededor de dichos valores en los últimos años. Otro aspecto importante es que, aunque los reactivos de dichos exámenes se han elaborado de manera colegiada, siempre han privilegiado el aspecto procedimental (centrándose en algoritmos y aplicación de fórmulas) y han ignorado sistemáticamente el aspecto conceptual de la materia (razonamiento inferencial e interpretación de resultados).

Así pues, el propósito de este trabajo es medir el razonamiento inferencial de los estudiantes de licenciatura del CUCEA de la U de G para diagnosticar e identificar fortalezas y áreas de oportunidad, ya que los futuros profesionales de esta área deben tomar decisiones importantes en el ámbito de empresas públicas o privadas, a partir de información recabada de muestras representativas, que deben utilizar para efectuar inferencias aplicables a sus poblaciones respectivas, con cierto nivel de confianza y margen de error. En este sentido, mientras más claros tengan los conceptos de la estadística inferencial, mayor certeza tendrán sus conclusiones, y eso se traducirá en decisiones con mayor probabilidad de éxito. Asimismo, al determinar el nivel de razonamiento inferencial y comprensión estadística de los estudiantes, permitirá impulsar cambios orientados al mejoramiento del proceso de enseñanza-aprendizaje de la estadística inferencial.

Para medir el razonamiento inferencial de los alumnos del CUCEA de la U de G se aplicó un instrumento de medición validado y calibrado, el CAOS-4, que además permite identificar las dificultades y errores específicos que presentan los alumnos, única y exclusivamente en cuanto al razonamiento inferencial. En este sentido, el razonamiento inferencial incluye una serie de saberes cognitivos relacionados con el área de la estadística inferencial, referentes a la comprensión conceptual de la disciplina, como el reconocimiento de distribuciones de probabilidad básicas y su representación gráfica, entendimiento de representaciones de relaciones bivariadas en diagramas de dispersión, identificación de patrones y variabilidad en muestras, la

¹ Los contenidos de estadística inferencial se imparten en una clase denominada Estadística II, existe también la materia de Estadística I, cuyos temas incluyen estadística descriptiva y probabilidad, cuya aplicación de exámenes departamentales también está a cargo de la misma Academia de Estadística.

capacidad de hacer interpretaciones razonables a partir de resultados obtenidos en pruebas de hipótesis o estimación de intervalos de confianza, entre otros (DelMas, Garfield, Ooms y Chance, 2007).

METODOLOGÍA

Para poder medir el nivel de razonamiento inferencial en estudiantes universitarios, específicamente de áreas económico-administrativas se aplica la versión en español del instrumento en inglés titulado *Comprehensive Assessment of Outcomes in Statistics* (CAOS) desarrollado por DelMas *et al.* (2007), el cual ha sido validado, calibrado y diseñado para estudiantes universitarios de los Estados Unidos de Norteamérica que han tomado un curso de estadística. El instrumento ha tenido varias modificaciones desde sus inicios, la versión que se ha considerado para esta investigación es la conocida como CAOS-4 (DelMas *et al.*, 2007). Dicho instrumento permite obtener información de cómo los estudiantes han comprendido los temas de estadística inferencial al completar su curso (DelMas *et al.*, 2007). La prueba CAOS-4 se desarrolló a través de un proceso de tres años de adquisición y redacción de ítems, contrastación y revisión de ítems, y recopilación de evidencia de confiabilidad y validez. Basado en una muestra de 23,645 estudiantes, un análisis de consistencia interna de los 40 ítems en la prueba CAOS-4 produjo un coeficiente alfa de Cronbach de 0.78 (DelMas *et al.*, 2007). Considerando lo anterior, se puede utilizar el instrumento CAOS-4 de manera confiable para medir el nivel de razonamiento estadístico en general y el razonamiento inferencial en estudiantes de licenciatura y con ello determinar el aprendizaje alcanzado por los alumnos en concordancia con los objetivos del curso de estadística inferencial.

El instrumento consta de cuarenta preguntas que incluyen temas de estadística descriptiva e inferencial, sin embargo, se tomaron solamente veintiuna preguntas que se refieren a temas de estadística inferencial: muestreo y distribuciones de muestreo, estimación de intervalos de confianza, pruebas de hipótesis y regresión lineal. Cada pregunta contiene una respuesta de entre dos y cuatro opciones. Se codificaron las respuestas dicotómicamente, en otras palabras, 1 si la respuesta es correcta y 0 si no lo es. Posteriormente, para presentar los resultados se transformaron en una escala de 0 a 100. Se capturaron en una base de datos las respuestas individuales de cada alumno por ítem del instrumento, para un análisis posterior más detallado por reactivo y por tema.

Cabe recalcar que la prueba CAOS-4 solo mide habilidades de razonamiento estadístico en general y de razonamiento inferencial en particular, no mide destrezas procedimentales tales como calcular, usar fórmulas o recordar definiciones textuales como las pruebas tradicionales, puesto que está diseñado para que los estudiantes piensen y razonen, no para que reproduzcan algoritmos y memoricen.

La muestra constó de 326 estudiantes que cursan la materia de estadística inferencial en tercero o cuarto semestre de las distintas licenciaturas económico-administrativas, cabe mencionar que tales alumnos aplicaron el instrumento de manera voluntaria y anónima (de una población de alrededor de 2,000 alumnos que fueron invitados a participar), para fines exclusivos de esta investigación. De los 326 alumnos considerados 192 eran mujeres y 134 eran hombres, las edades de los alumnos oscilaban entre 19 y 27 años, e incluía tanto alumnos del turno matutino como del turno vespertino.

El instrumento se implementó al final del primer semestre del 2018, cuando concluyeron su materia de estadística inferencial. Se aplicó de manera simultánea a todos los estudiantes con el apoyo de un grupo de profesores; el tiempo máximo que tuvieron los alumnos para finalizar la prueba fue de 45 minutos; los alumnos no podían utilizar la calculadora ya que se enfatizó el carácter de razonamiento inferencial del examen.

Con los resultados individuales de las 326 pruebas aplicadas se efectuarán los siguientes procedimientos: a) el cálculo de estadísticos descriptivos generales de la prueba aplicada; b) el análisis por reactivo y por tema del instrumento, que considera tanto el aprendizaje esperado y evaluado como las opciones de respuesta con la justificación de los distractores, y c) la obtención de la calificación promedio por temas, y el análisis del razonamiento inferencial por temas.

RESULTADOS

A partir de los resultados del instrumento CAOS-4 se observa que el nivel de aprendizaje de los estudiantes del CUCEA en los temas de estadística inferencial muestra marcadas deficiencias. La tabla 1 muestra las estadísticas descriptivas de los resultados

Tabla 1. Estadísticos descriptivos.

Estadísticos	Valores
Media	39.33
Mediana	38.10
Moda	47.62
Desviación estándar	11.04
Varianza de la muestra	121.88
Curtosis	-0.50
Coficiente de asimetría	-0.01
Rango	57.14
Mínimo	14.29
Máximo	71.43

Fuente: Elaboración propia.

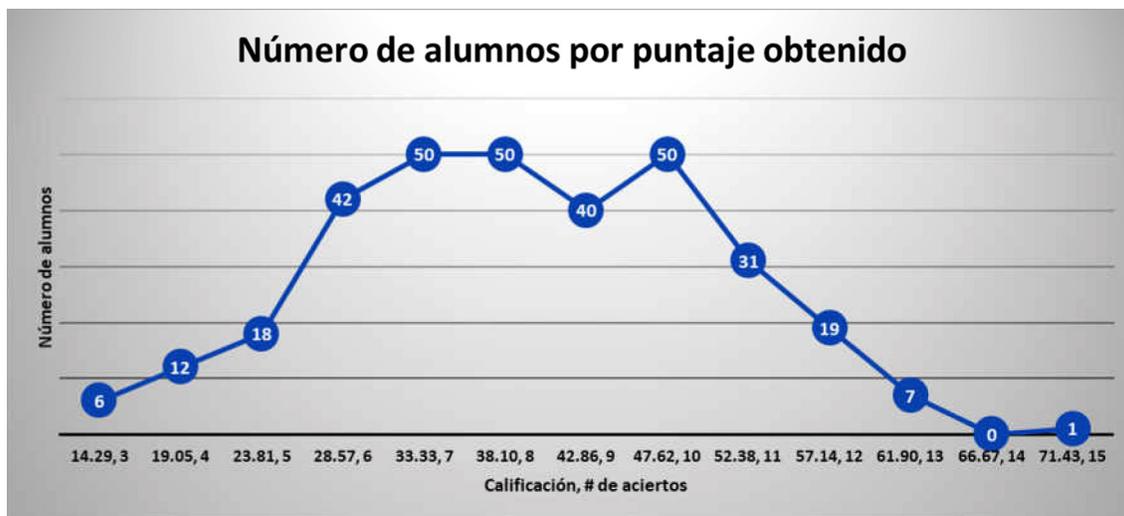


Figura 1. Número de alumnos por calificación obtenida en la prueba CAOS-4.

Fuente: Elaboración propia.

del instrumento aplicado en una escala de 0 a 100, que es una escala más afín a la manera de evaluar a los alumnos en las asignaturas universitarias. El valor mínimo obtenido fue de 14.28 (3 respuestas correctas de 21 preguntas) y el máximo de 71.43 (15 respuestas correctas). No hubo ningún estudiante que tuviera todas las respuestas correctas. Se observa que el promedio de calificación es de 39.33, dicho promedio es más pequeño que el promedio obtenido en los años recientes de aplicación de exámenes departamentales (que oscila entre 45 y 49 puntos), lo cual muestra que el nivel de razonamiento inferencial obtenido mediante el instrumento CAOS-4 es menor que el aprendizaje de tipo procedimental basado en algoritmos y aplicación de fórmulas medido a través de los exámenes departamentales de estadística inferencial.

Al mismo tiempo, la mediana y la moda fueron respectivamente 38.10 y 47.62. La distribución de los datos es asimétrica negativa. Esto demuestra que existen más datos por debajo de la media y la forma de la distribución es mesocúrtica, así que muestra que hay una gran cantidad de datos alrededor de la mediana (ver figura 1).

El promedio por temas en una escala de 0 a 100 (ver figura 2) se muestra a continuación: muestreo y distribuciones de muestreo: 41 (reactivos 1, 2, 17, 18 y 19), estimación de intervalos de confianza: 42 (reactivos 12, 13, 14, y 15), prueba de hipótesis: 35 (reactivos 3, 7, 8, 9, 10, 11, 16 y 21) y regresión simple: 44 (reactivos 4, 5, 6 y 20). De los 326 alumnos que respondieron el instrumento solo 8 obtuvieron una calificación aprobatoria (mayor o igual que 60). Aunque no obstante que los cuatro temas de estadística inferencial presentan dificultades de razonamiento, el tema de prueba de hipótesis tiene mayor dificultad respecto a los demás.

Se realiza un análisis por reactivo del instrumento y por tema, que está basado en la rúbrica de los reactivos de la prueba CAOS-4 (mostrados en las tablas 2, 3, 4 y



Figura 2. Promedio por tema en la prueba CAOS4.

Fuente: Elaboración propia.

5), el cual contiene la descripción del aprendizaje esperado y evaluado, las opciones de respuesta con la justificación de los distractores; en negritas se señala la opción correcta, así como los porcentajes correspondientes a cada inciso, que se obtuvieron de los resultados de la aplicación de la prueba CAOS-4 a la muestra elegida. Cabe mencionar que todos los indicadores de las tablas 2, 3, 4 y 5 se tomaron del instrumento CAOS-4.

En lo que respecta a los reactivos de muestreo se puede mencionar lo siguiente (ver tabla 2): para el reactivo 1, solo el 29% de los alumnos comprendió que si se tienen varias muestras aleatorias del mismo tamaño de cierta población se puede esperar un patrón predecible con algunas variaciones entre las muestras. En lo referente al reactivo 2, el 67% de los estudiantes entendió bien el significado de la variabilidad en el contexto de intentos repetidos en una situación donde se espera que esta sea pequeña. En lo que se refiere al reactivo 17, el 56% de los alumnos comprendió correctamente que una muestra grande explica mejor el comportamiento de una población que una muestra pequeña. En cuanto al reactivo 18, solo el 27% de los estudiantes tuvo la habilidad de seleccionar una distribución de muestreo apropiada para una población particular de acuerdo al tamaño de muestra dado. Para el reactivo 19, el 25% de los alumnos comprendió adecuadamente que a partir de una muestra aleatoria de tamaño razonable se pueden obtener resultados confiables que se pueden generalizar a toda la población.

Entre los errores más comunes del tema de muestreo se pueden mencionar los siguientes: creer que un patrón idéntico debe repetirse necesariamente en cada muestra, no identificar la diferencia en la dispersión entre dos conjuntos de valores, creer que la distribución de los datos de una muestra grande tiene que ser normal, suponer que la distribución de muestreo y la distribución poblacional deben ser similares, no identificar una muestra representativa o un tamaño de muestra adecuado.

Tabla 2. Rúbrica de los reactivos del tema distribuciones de muestreo de la prueba CAOS-4.

Ítem	Resultado del aprendizaje evaluado	Opción de respuesta %			
		a)	b)	c)	d)
1	Comprender que si se tienen varias muestras aleatorias del mismo tamaño de cierta población dada la variación de una muestra a otra puede identificar que seguirá un patrón predecible	(29%) Patrón esperado	(27%) Patrón repetido (poco probable)	(8%) Patrón discordante con la porción poblacional	(36%) Razonamiento incorrecto
2	Comprender el significado de variabilidad en el contexto de intentos repetidos y en una situación donde se desea una pequeña variabilidad	(67%) El conjunto de datos es más homogéneo (menor s)	(18%) El conjunto de datos es más heterogéneo	(15%) Razonamiento incorrecto	
17	Comprender que una muestra grande explica de manera más apropiada el comportamiento de una población que una muestra pequeña	(56%) La distribución muestral se parece a la distribución poblacional para un tamaño de muestra grande	(33%) La distribución de la muestra se parece a la normal, mientras que la distribución poblacional no	(11%) La distribución de la muestra es muy diferente de la distribución poblacional	
18	Habilidad de seleccionar una distribución de muestreo apropiada para una población particular de acuerdo al tamaño de muestra dado	(24%) La distribución de muestreo no se aproxima a la normal	(27%) La distribución de muestreo se aproxima a la normal con $\mu=\mu_x$	(49%) Claramente $\mu>\mu_x$.	
19	Comprender los factores que permiten que los datos obtenidos de una muestra se puedan generalizar a la población	(25%) La muestra es aleatoria y el tamaño de la misma adecuado	(22%) La muestra no es representativa	(15%) El tamaño de muestra puede no ser adecuado o la muestra puede estar sesgada	(38%) Razonamiento incorrecto

Fuente: Elaboración propia basada en el instrumento CAOS-4 elaborado por DelMas, Garfield, Ooms y Chance (2007).

En lo concerniente a estimaciones de intervalo, se encontraron los siguientes hallazgos (ver tabla 3): para los reactivos 12, 13, 14 y 15 se construyó un intervalo de confianza del 95% con los datos de una muestra y se hicieron cuestionamientos de falso o verdadero. En referencia al reactivo 12, solo el 28% de los estudiantes detectó que es erróneo afirmar que cada media muestral está en el intervalo de confianza con una probabilidad del 95%. Con respecto al reactivo 13, solo 44% reconoció que es equivocado suponer que el 95% de las observaciones muestrales está en el intervalo de confianza. En cuanto al reactivo 14, solamente el 41% de los alumnos advirtió que es erróneo pensar que el 95% de todas las medias muestrales está en el intervalo de confianza. Y referente al reactivo 15, el 61% de los estudiantes interpretó

Tabla 3. Rúbrica de los reactivos del tema intervalos de confianza de la prueba CAOS-4.

Ítem	Resultado del aprendizaje evaluado	Opción de respuesta %			
		a)	b)	c)	d)
12	Habilidad para detectar una mala interpretación del nivel de confianza (el porcentaje muestral está entre los límites de confianza)	(72%) Válido	(28%) Inválido		
13	Habilidad para detectar una mala interpretación del nivel de confianza (el porcentaje poblacional está entre los límites de confianza)	(56%) Válido	(44%) Inválido		
14	Habilidad para entender una mala interpretación del nivel de confianza (el promedio de los porcentajes muestrales está entre los límites de confianza)	(59%) Válido	(41%) Inválido		
15	Habilidad para interpretar correctamente un intervalo de confianza	(61%) Válido	(39%) Inválido		

Fuente: Elaboración propia basada en el instrumento CAOS-4 elaborado por DelMas, Garfield, Ooms y Chance (2007).

correctamente el intervalo de confianza (con una probabilidad del 95% el intervalo de confianza incluye la media aritmética poblacional).

Los errores más frecuentes del tema de intervalos de confianza consisten en interpretaciones erróneas del nivel de confianza, como asumir el nivel de confianza como el porcentaje de datos de la población que está en el intervalo de confianza, o de datos de cualquier muestra que están en el intervalo de confianza, o como el porcentaje de medias muestrales que está en el intervalo de confianza, o que dicho nivel de confianza representa la probabilidad de que cualquier media muestral esté contenida en el intervalo de confianza.

Por otro lado, en cuanto a los reactivos referentes a prueba de hipótesis se destacan las siguientes consideraciones (ver tabla 4): en referencia el reactivo 3, el 24% de los estudiantes advirtió correctamente que los valores p pequeños son deseables en estudios de investigación. Respecto al reactivo 7, el 49% de los alumnos respondió correctamente al concluir que, aunque no se rechace la hipótesis nula, no se puede descartar la hipótesis alternativa. En lo concerniente al reactivo 8, el 57% comprendió de manera correcta que rechazar la hipótesis nula implica que es muy probable que la hipótesis alternativa sea cierta. Para los reactivos 9, 10 y 11 se planteó una prueba de hipótesis y se formularon preguntadas de válido e inválido acerca de la interpretación del valor p pequeño. En el reactivo 9, el 45% de los estudiantes entendió que un valor p pequeño muestra que la hipótesis nula es muy improbable que ocurra; para el reactivo 10, solo el 34% de los alumnos comprendió que dicho valor p , no representa la probabilidad de que la hipótesis nula ocurra; mientras para el reactivo 11,

Tabla 4. Rúbrica de los reactivos del tema prueba de hipótesis del instrumento CAOS-4.

Ítem	Resultado del aprendizaje evaluado	Opción de respuesta %			
		a)	b)	c)	d)
3	Comprender que los valores pequeños del valor de p son deseables en estudios de investigación en pruebas de significancia	(34%) El valor p no muestra significancia estadística	(24%) Correcto	(42%) Razonamiento incorrecto	
7	Comprender que, aunque no haya significancia estadística eso no garantiza que se pueda descartar la hipótesis alternativa (H_a)	(13%) El hecho que no haya significancia estadística no garantiza que se pueda descartar H_a	(49%) El tamaño de muestra puede ser muy pequeño para detectar una diferencia significativa	(38%) No rechazar la hipótesis nula (H_0) no significa necesariamente que sea cierta	
8	Comprender que de un diseño experimental con asignación aleatoria se puede inferir causalidad	(19%) La prueba de hipótesis no demuestra la asociación entre variables	(9%) Se puede tomar el estadístico conveniente para el tamaño de muestra aunque sea pequeño	(15%) La prueba de hipótesis no establece relación lineal entre variables	(57%) Rechazar H_0, implica que H_a sea muy probable que ocurra
9	Habilidad para reconocer correctamente la interpretación de un valor p en pruebas de significancia	(45%) Válido	(55%) Inválido		
10	Habilidad para reconocer incorrectamente la interpretación de un valor p . Específicamente, la probabilidad de que un tratamiento no sea efectivo	(66%) Válido	(34%) Inválido		
11	Habilidad para reconocer incorrectamente la interpretación de un valor p . Específicamente, la probabilidad de que un tratamiento sea efectivo	(66%) Válido	(34%) Inválido		
16	Comprender como el error de muestreo se utiliza para efectuar una inferencia informal sobre una media muestral	(46%) No se está considerando el tamaño de la muestra	(38%) No se está considerando el error de muestreo	(16%) Se puede utilizar el error de muestreo para establecer la conclusión informal	
21	Comprender la lógica de una prueba de significancia cuando se rechaza la hipótesis nula	(50%) Deducir que rechazar H_0 implica que sea necesariamente falsa	(21%) Rechazar H_0 no implica necesariamente que sea falsa	(14%) Rechazar H_0 pero aceptar su implicación de manera tajante	(15%) Rechazar H_0 pero aceptar su implicación de manera moderada

Fuente: Elaboración propia basada en el instrumento CAOS-4 elaborado por DelMas, Garfield, Ooms y Chance (2007).

Tabla 5. Rúbrica de los reactivos del tema regresión simple de la prueba CAOS-4.

Ítem	Resultado del aprendizaje evaluado	Opción de respuesta %			
		a)	b)	c)	d)
4	Habilidad para hacer coincidir la información de un diagrama de dispersión con una descripción verbal en datos bivariados	(81%) Muestra una relación lineal negativa	(14%) Muestra una relación lineal directa	(5%) El diagrama no muestra relación entre las variables	
5	Habilidad para entender la relación bivariada que existe en un diagrama de dispersión cuando existe un valor atípico	(25%) Considerar el valor atípico aparenta una fuerte relación lineal negativa entre las variables	(21%) Considerar el valor atípico muestra una aparente relación lineal entre las variables	(54%) No hay relación entre las variables. El valor atípico se puede ignorar	
6	Comprender que la correlación no implica causalidad en la recopilación y diseño de datos	(24%) Correcta	(26%) El tamaño de muestra no es relevante en este caso	(50%) Concluir que la correlación implica causalidad	
20	Comprender cuando no es recomendable extrapolar usando un modelo de regresión	(28%) El año pronosticado es muy posterior al periodo utilizado en el modelo de regresión	(28%) El año pronosticado es muy posterior al periodo utilizado en el modelo de regresión	(14%) El modelo de regresión no considera variaciones cíclicas ni estacionales	(30%) Razonamiento incorrecto

Fuente: Elaboración propia basada en el instrumento CAOS-4 elaborado por DelMas, Garfield, Ooms y Chance (2007).

solamente el 34% de los estudiantes opinó acertadamente que tal valor p no representa la probabilidad de que ocurra la hipótesis alternativa. Por lo que toca al reactivo 16, referente a una prueba de hipótesis de una muestra para medias, únicamente el 16% de los alumnos dedujo de forma correcta que se puede utilizar el error de muestreo para establecer una conclusión informal en una prueba de hipótesis. Y finalmente, en relación al reactivo 21, solamente el 21% de los estudiantes concluyó apropiadamente que rechazar la hipótesis nula no implica que necesariamente sea falsa.

Con respecto al tema de prueba de hipótesis los errores más frecuentes son: suponer que cuando no se rechaza la hipótesis nula esta debe ser necesariamente cierta, o que cuando se rechaza la hipótesis nula, la hipótesis alternativa se cumple de manera forzosa; creer que los valores p grandes son los más convenientes en estudios de investigación, o que dicho valor no tiene relevancia en la significancia estadística; presumir que para efectuar una prueba de hipótesis basta comparar la media supuesta y la media muestral, sin considerar la desviación estándar y el tamaño de la muestra.

Ahora bien, para los reactivos referentes al tema de modelo de regresión lineal se observa lo siguiente (ver tabla 5): para el reactivo 4, el 81% de los alumnos interpretó correctamente la información del diagrama de dispersión como una relación lineal

entre dos variables. Respecto al reactivo 5, el 54% de los estudiantes respondió correctamente al no identificar ninguna relación lineal entre las variables en un diagrama de dispersión, no obstante la presencia de un valor atípico. En lo que concierne al reactivo 6, el 24% de los alumnos comprendió adecuadamente que la correlación no implica causalidad. Finalmente, en relación al reactivo 20, solo el 14% de los estudiantes contestó correctamente al entender que el modelo de regresión no puede utilizarse para efectuar un pronóstico en un horizonte de tiempo lejano, sin considerar las variaciones comunes de las series temporales.

En cuanto al tema de regresión, entre los errores más comunes destacan los siguientes: confundir entre relaciones lineales positivas y negativas a partir de un diagrama de dispersión; suponer que la correlación entre dos variables implica causalidad; creer que se pueden efectuar pronósticos en el largo plazo, tanto gráficamente a partir de la línea de regresión como analíticamente a través de la ecuación de regresión, sin considerar las variaciones típicas de las series de tiempo.

CONCLUSIONES

Los resultados del instrumento CAOS-4 muestran que el nivel de razonamiento inferencial de los estudiantes del CUCEA en todos los temas de estadística inferencial no está siendo el adecuado. En este sentido, podemos mencionar que, de los 21 reactivos del instrumento CAOS-4 para medir habilidades de razonamiento inferencial, solo en 3 de ellos mostraron un porcentaje aprobatorio de respuestas correctas (superior a 60%), estas conductas fueron: comprender el significado de variabilidad en el contexto de intentos repetidos, interpretar correctamente un intervalo de confianza como el rango de valores en el que es posible que se encuentre un parámetro poblacional e identificar la información de un diagrama de dispersión con una descripción verbal de datos bivariados. Por otro lado, en el tema de prueba de hipótesis se obtuvo un promedio significativamente menor a los temas de muestreo y distribuciones de muestreo, estimaciones de intervalos de confianza y regresión lineal.

El instrumento aplicado prioriza la parte conceptual sobre la procedimental. En este sentido, se identificaron algunas deficiencias de razonamiento inferencial notorias de los estudiantes en el área de estadística inferencial, entre las que podemos mencionar las siguientes: no entienden el carácter aleatorio de las muestras como subconjunto de una población con variabilidad implícita, lo cual significa que las conclusiones derivadas de las muestras aplicables a la población tienen un carácter probabilista y no determinista; no comprenden la importancia del tamaño de muestra como indicador de la representatividad, y como indicador de la confianza en las inferencias poblacionales; confunden las estadísticas muestrales y los parámetros poblacionales;

tienen dificultades para relacionar el nivel de confianza y el margen de error en las estimaciones de intervalo de confianza de la media; así como en la interpretación de dichos intervalos o la utilización errónea de la distribución de probabilidad adecuada de acuerdo al tamaño de muestra en la construcción de tales intervalos de confianza; en las pruebas de hipótesis no establecen las hipótesis adecuadas, no distinguen claramente entre hipótesis nula e hipótesis alternativa, así como dificultades para elegir un adecuado nivel de significancia (y la relación de este con el valor p en la contrastación de la hipótesis); confunden la hipótesis nula con lo que se desea probar (no lo que se desea rechazar); ignoran que, ya sea que se rechace o no la hipótesis nula, la conclusión es susceptible de error, ya sea de rechazar la hipótesis nula cuando es cierta o de aceptar dicha hipótesis nula cuando es falsa con una probabilidad asociada en cada caso; no identifican una relación lineal entre dos variables dado un diagrama de dispersión, o no reconocen cuando una relación lineal es positiva o negativa; intentan aplicar la ecuación de regresión a valores de la variable independiente fuera del rango en el que se encuentran las variables; del mismo modo, tratan de aplicar el método de los mínimos cuadrados cuando la relación entre las variables no es lineal; entre otros errores comunes.

Por otro lado, al ser los estudiantes de la muestra a la que se aplicó el CAOS-4 (un test de razonamiento inferencial) un subconjunto del total de alumnos a los que se aplicó el examen departamental de estadística inferencial en el primer semestre del año 2018 (un test de tipo procedimental), podemos comparar los promedios 39.33 versus 47.31 y afirmar que en ambos casos los alumnos muestran series deficiencias, aunque es significativamente menor el nivel de razonamiento inferencial que el nivel de habilidades meramente procedimentales. Los resultados obtenidos concluyen que los alumnos no están alcanzando los niveles óptimos en cuanto a las habilidades de razonamiento deseadas en la materia de estadística inferencial. En este sentido, se recomienda que los cursos de estadística inferencial enfatizen el aspecto conceptual y la lógica subyacente a las pruebas y procedimientos, de la misma forma en que se hace hincapié en las fórmulas y algoritmos de tal manera que los estudiantes sepan aplicar tales metodologías al tiempo que desarrollan su razonamiento inferencial, conscientes de las limitaciones y del carácter aleatorio de la disciplina.

De acuerdo a los resultados de este estudio, y como futuras líneas de investigación para descubrir las causas del bajo nivel de los estudiantes en cuanto al razonamiento inferencial, se pueden mencionar algunas hipótesis: la deficiente capacitación didáctica de los profesores que imparten la materia de estadística inferencial; alumnos con mala preparación matemática y estadística; nulo o escaso uso de *software* estadístico, bases de datos reales y estudios de casos prácticos.

REFERENCIAS

- AME (2020). *Asociación Mexicana de Estadística*. Recuperado de: <http://amestad.mx>.
- Beitz, J. (1998). Helping students learn and apply statistical analysis: A metacognitive approach. *Nurse Education*, 23(1), 49-51. <https://doi.org/10.1097/00006223-199801000-00016>.
- Cuevas, J. H., e Ibáñez, C. (2008). Estándares en educación estadística: necesidad de conocer la base teórica y empírica que los sustentan. *Revista Iberoamericana de Educación Matemática*, (15), 33-45. Recuperado de: <http://funes.uniandes.edu.co/14850/1/Cuevas2008Est%C3%A1ndares.pdf>.
- DelMas, R., Garfield, J. B., Ooms, A., y Chance, B. (2007). Assessing students' conceptual understanding after a first course in Statistics. *Statistics Education Research Journal*, 6(2), 28-58. <https://doi.org/10.1016/j.brat.2010.07.005>.
- Delucchi, M. (2014). Measuring student learning in social statistics: A pretest-posttest study of knowledge gain. *Teaching Sociology*, 42(3), 231-239. <https://doi.org/10.1177/0092055X14527909>.
- Doerr, H. M., Delmas, R., y Makar, K. (2017). A modeling approach to the development of students' informal inferential reasoning. *Statistics Education Research Journal*, 16(2), 86-115. Recuperado de: http://iase-web.org/documents/SERJ/SERJ16%282%29_Doerr.pdf.
- Gibbs, B. G., Shafer, K., y Miles, A. (2017). Inferential statistics and the use of administrative data in US educational research. *International Journal of Research and Method in Education*, 40(2), 214-220. <https://doi.org/10.1080/1743727X.2015.1113249>.
- Huang, Z. X. (2018). Application of statistical inference in education and teaching. *Educational Sciences: Theory and Practice*, 18(6), 2782-2793. <https://doi.org/10.12738/estp.2018.6.179>.
- Hubbard, R., Haig, B. D., y Parsa, R. A. (2019). The limited role of formal statistical inference in scientific inference. *The American Statistician*, 73(sup. 1), 91-98. <https://doi.org/10.1080/00031305.2018.1464947>.
- Makar, K., y Rubin, A. (2017). Learning about statistical inference. En D. Ben-Zvi, K. Makar y J. B. Garfield (eds.), *International handbook of research in Statistics education* (pp. 261-294). Cham, Suiza: Springer. https://doi.org/10.1007/978-3-319-66195-7_8.
- Noll, J., Gebresenbet, M., y Glover, E. D. (2016). A modeling and simulation approach to informal inference: Successes and challenges. *The Teaching and Learning of Statistics*, 2016(1), 139-150. https://doi.org/10.1007/978-3-319-23470-0_19.
- Peñaloza, J. L., y Vargas, C. (2017). Big-data and the challenges for statistical inference and economics teaching and learning. *Multidisciplinary Journal for Education, Social and Technological Sciences*, 4(1), 64-87. <https://doi.org/10.4995/muse.2017.6350>.
- Pfannkuch, M., y Wild, C. J. (2015). Laying foundations for statistical inference. En S. J. Cho (ed.), *Selected regular lectures from the 12th International Congress on Mathematical Education* (pp. 653-666). Cham, Suiza: Springer. https://doi.org/10.1007/978-3-319-17187-6_36.
- Reaburn, R. (2018). Students' understanding of statistical inference: Implications for teaching. En D. Kember y M. Corbett (eds.), *Structuring the thesis* (pp. 121-127). Singapur: Springer. https://doi.org/10.1007/978-981-13-0511-5_12.
- Rumsey, D. J. (2002). Statistical literacy as a goal for introductory Statistics courses. *Journal of Statistics Education*, 10(3), 1-12. <https://doi.org/10.1080/10691898.2002.11910678>.
- Sánchez, J. (2010). *International statistical literacy project*. Recuperado de: <http://www.stat.fi/org/tilastokeskus/flyer.pdf>.
- Tong, C. (2019). Statistical inference enables bad science; statistical thinking enables good science. *The American Statistician*, 73(1), 246-261. <https://doi.org/10.1080/0031305.2018.1518264>.
- Watson, J. M. (1997). Assessing statistical thinking using the media. En E. Gal y J. B. Garfield (eds.), *The assessment challenge in Statistics education* (pp. 107-121). Amsterdam, Holanda: IOS Press. Recuperado de: <http://iase-web.org/documents/book1/chapter09.pdf>.
- Weinberg, A., Wiesner, E., y Pfaff, T. J. (2010). Using informal inferential reasoning to develop formal concepts: Analyzing an activity. *Journal of Statistics Education*, 18(2), 1-24. <https://doi.org/10.1080/10691898.2010.11889494>.
- Zellner, K., Boerst, C. J., y Tabb, W. (2007). Statistics used in current nursing research. *Journal Nurse Education*, 46(2), 55-59. Recuperado de: <https://pdfs.semanticscholar.org/0a00/93cc3d131534073ee13bc3bafb014ac2b46a.pdf>.

Cómo citar este artículo:

Sandoval Bravo, S., Celso-Arellano, P. L., y Gualajara Estrada, V. H. (2021). Diagnóstico sobre el razonamiento inferencial en estudiantes de licenciatura del área económico-administrativa. *IE Revista de Investigación Educativa de la REDIECH*, 12, e1134. doi: 10.33010/ie_rie_rediech.v12i0.1134.



Todos los contenidos de *IE Revista de Investigación Educativa de la REDIECH* se publican bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional, y pueden ser usados gratuitamente para fines no comerciales, dando los créditos a los autores y a la revista, como lo establece la licencia.