

# Dificultades de comprensión del intervalo de confianza en estudiantes universitarios

Eusebio Olivo, Carmen Batanero y Carmen Díaz

**Resumen:** En este trabajo presentamos un estudio de evaluación de la comprensión de estudiantes universitarios de ingeniería sobre los intervalos de confianza. Un cuestionario formado por 12 ítems de opción múltiple y dos problemas abiertos fue respondido por una muestra de 252 estudiantes de ingeniería que habían estudiado el tema. Los resultados indican que los estudiantes no comprenden el carácter aleatorio de los extremos de los intervalos, asignan una interpretación bayesiana al coeficiente de confianza o no comprenden el efecto de varios factores sobre la anchura del intervalo. Otras dificultades son la selección de una distribución adecuada al construir un intervalo, el cálculo de los grados de libertad o la determinación de valores críticos.

*Palabras clave:* intervalo de confianza, evaluación, estudiantes universitarios.

**Abstract:** We present the results from a research directed to assess students' understanding of confidence intervals in engineering undergraduate students. A questionnaire composed by 12 multiple choice items and two open problems was given to a sample of 252 engineering students after they studied the topic. Results suggest that students do not understand the random character of the interval limits, assign a Bayesian interpretation to the confidence coefficient or do not understand the effect of several factors on the interval width. Other difficulties are related to selecting an adequate distribution in building an interval, computing the degrees of freedom or finding the critical values.

*Keywords:* confidence intervals, assessment, university students.

---

Fecha de recepción: 24 de septiembre de 2007.

## INTRODUCCIÓN

El intervalo de confianza es un tema estudiado en todos los cursos universitarios de estadística e incluso en la educación preuniversitaria en países como España, donde se incluye en el Bachillerato de Ciencias Sociales. Es también base de muchos métodos estadísticos, tales como el diseño de experimentos, la regresión, estimación y métodos no paramétricos. Por otro lado, diferentes asociaciones profesionales, como la American Psychological Association (APA) o la American Educational Research Association (AERA) han denunciado los errores frecuentes en el uso de la inferencia por parte de los investigadores (Morrison y Henkel, 1970; Vallecillos, 1994; Harlow, Mulaik y Steiger, 1997; Batanero, 2000). Estos mismos autores sugieren que los intervalos de confianza podrían sustituir o complementar los contrastes de hipótesis para paliar los problemas citados.

Este cambio metodológico requiere asegurar que las dificultades descritas en la interpretación de los tests de hipótesis no se repiten –o al menos no con tanta intensidad– en los intervalos de confianza, tema donde la investigación didáctica es todavía incipiente. Por ello, resulta importante llevar a cabo estudios de evaluación sobre esta comprensión, que pudiera no ser inmediata, puesto que el intervalo de confianza se apoya en otros conocimientos previos y es un objeto matemático complejo al involucrar diferentes conceptos, procedimientos y propiedades.

En el contexto de estimar un parámetro poblacional, un intervalo de confianza se define como un rango de valores que depende de un coeficiente de confianza. Este coeficiente indica el porcentaje de muestras tomadas en las mismas condiciones, en las cuales el intervalo cubriría el verdadero valor del parámetro.

Hay un procedimiento general (Devore, 2005) para la construcción de dicho rango de valores a partir de un estadístico, calculado en los datos de la muestra, que se usa como estimador del parámetro correspondiente. Calculada la desviación típica del estadístico (error estándar) y obtenido un valor crítico correspondiente a la mitad del valor del coeficiente de confianza elegido, el producto del valor crítico por el error estándar se sumaría y restaría al valor del estadístico en la muestra, obteniendo así los límites del intervalo.

Este procedimiento general se particulariza dependiendo del parámetro que se va a estimar (media, proporción, varianza, etc.) y según las condiciones (tipo de distribución, qué se conoce de la misma, etc.), puesto que ellas determinan la distribución muestral del estadístico. Vemos, entonces, que la comprensión del

intervalo de confianza requiere una serie de otros objetos matemáticos previos (tanto conceptos como procedimientos), como población y muestra, estadístico y parámetro, error estándar y cálculo de éste para diversos estadísticos, distribución muestral, valor crítico o uso de las tablas de diferentes distribuciones.

## MARCO TEÓRICO

En el modelo teórico que nos sirve de base (Godino y Batanero, 1998; Godino, 2002), el objeto matemático se concibe emergiendo progresivamente del sistema de prácticas socialmente compartidas, ligadas a la resolución de cierto campo o tipo de problemas matemáticos. El *significado del objeto* sería una entidad compuesta formada por el conjunto de prácticas operatorias y discursivas relacionado con dicho campo de problemas. En estas prácticas intervienen como entidades primarias no sólo las definiciones del objeto, sino sus propiedades, problemas asociados, procedimientos, representaciones y argumentos en los que interviene. En nuestro caso, el objeto intervalo de confianza surge de las prácticas matemáticas relacionadas con la solución de los problemas de estimación y comparación de parámetros.

Los autores diferencian entre *significado institucional y personal* del objeto dado, según que las prácticas relativas a un cierto campo de problemas sean compartidas dentro de una institución (en nuestro caso la institución de enseñanza donde se tomaron los datos) o sean específicas de un sujeto de dicha institución (los alumnos participantes en el estudio).

Desde el punto de vista de la institución, un sujeto “comprende” el significado de un objeto si es capaz de realizar correctamente las distintas prácticas que configuran el significado de dicho objeto institucional. La comprensión de un objeto como el intervalo de confianza no será, por tanto, unitaria, sino que abarca cada uno de los tipos de entidades primarias consideradas y se construye progresivamente en el proceso de aprendizaje.

En nuestra investigación preparamos un cuestionario comprensivo de evaluación de las dificultades de comprensión del intervalo de confianza, incluidas las diferentes entidades primarias que lo conforman según el marco teórico. Nos basamos también en las investigaciones previas que se describen a continuación.

## INVESTIGACIONES SOBRE COMPRENSIÓN DE LOS INTERVALOS DE CONFIANZA

Las investigaciones previas sobre comprensión del intervalo de confianza son pocas. Entre las que se centran en investigadores, destacamos la de Cumming, William, y Fidler (2004), que estudian los errores en la interpretación de intervalos de confianza de la media de una población, sugiriendo que se espera (erróneamente) una alta probabilidad de replicación, suponiendo que, al tomar una nueva muestra, el nuevo intervalo de confianza será muy parecido al original. Según Schenker y Gentleman (2001), otra creencia errónea común en los investigadores es que los intervalos de confianza de dos medias de muestras independientes son sólo significativamente diferentes cuando se tocan justo extremo con extremo. También confunden el cálculo de intervalos de confianza para medias independientes y relacionadas.

Respecto a los trabajos con estudiantes universitarios, Fidler y Cumming (2005) indican que 44% de los estudiantes interpreta incorrectamente un *valor p* pequeño en un contraste de hipótesis, pero sólo 18% interpreta incorrectamente los resultados cuando se presentan mediante intervalos de confianza, lo que sugiere que la comprensión de los intervalos de confianza podría ser más sencilla que la relacionada con el contraste de hipótesis. Otros estudiantes en la investigación de Fidler y Cumming sólo consideran los intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial o tienen ideas equivocadas sobre cómo se relacionan entre sí los distintos conceptos que intervienen en los intervalos de confianza. Por ejemplo, sólo 16% de los estudiantes en la investigación citada pudo contestar correctamente a la relación entre ancho del intervalo y tamaño de la muestra.

Garfield, delMas y Chance (1999), usando un *software* de simulación de elaboración propia, tratan de favorecer el aprendizaje de conceptos básicos de la estadística, entre ellos los intervalos de confianza. Los autores plantean un modelo en el cual destacan los elementos conceptuales que el estudiante debería entender acerca de los intervalos de confianza, las competencias básicas que el estudiante debería adquirir y algunas concepciones erróneas.

Basándose en dicho trabajo, Behar (2001) construye un cuestionario sobre comprensión del intervalo de confianza y contraste de hipótesis. De las respuestas al cuestionario, concluye que los alumnos no relacionan la influencia del nivel de confianza en el ancho del intervalo. También observa dificultad de comprensión en la definición del intervalo de confianza, porque se piensa que los

valores que lo constituyen son los que toma la variable aleatoria que define la población o valores del estadístico que se usa como estimador; lo cual podría deberse al propio procedimiento de construcción del intervalo, el cual se deduce de la distribución muestral de dicho estadístico. Los participantes en su estudio no asocian el nivel de confianza con la frecuencia relativa con la que los intervalos generados por un mecanismo aleatorio de obtención de muestras repetidas de la misma población incluyen el verdadero parámetro de la población. La utilidad de los intervalos de confianza para tomar decisiones sobre hipótesis parece no ser comprendida, posiblemente por no considerar los valores del intervalo como un conjunto de valores plausibles del parámetro.

Terán (2006) investiga el significado de los intervalos de confianza para estudiantes del primer año de un curso de estadística en la Universidad Nacional de Rosario, Argentina. Basándose en el modelo teórico de la cognición matemática de Godino (2002), su propósito es detectar cómo se construyen los diferentes elementos de significado (definición, propiedades, argumentos, problemas, representaciones) del intervalo de confianza en dos estudiantes, a través de su interacción con el profesor y la computadora cuando resuelven un problema. La autora observa cómo la interacción con la computadora incrementa la relevancia del aprendizaje, pues los estudiantes exploran y experimentan diferentes situaciones que dan sentido a dichos elementos de significado.

En este trabajo continuamos las investigaciones anteriores, presentando los resultados obtenidos a través de las respuestas a un cuestionario de 252 estudiantes de ingeniería en México. El objetivo principal es hacer un diagnóstico sobre la comprensión que de los intervalos de confianza tienen los alumnos universitarios, en las diferentes entidades primarias consideradas en nuestro marco teórico. Por otro lado, la mayoría de sus ítems podrían aplicarse en el bachillerato en aquellas especialidades que incluyen el tema. Por tanto, las dificultades descritas en este trabajo podrían orientar la labor del profesor en cualquiera de los dos niveles educativos.

## METODOLOGÍA DEL TRABAJO

El primer paso en la construcción del cuestionario fue llevar a cabo una definición precisa de la variable objeto de medición. Puesto que la *comprensión del intervalo de confianza* es un constructo psicológico inobservable, es preciso llevar a cabo una definición detallada de los contenidos que se evaluarán. Nuestra

finalidad es analizar la comprensión de los intervalos de confianza cuando se considera “comprensión” en este sentido global y en la institución considerada.

Ya que el estudio se centra en estudiantes de ingeniería, se llevó a cabo un análisis del contenido relacionado con los intervalos de confianza en una muestra de 11 libros de estadística dirigidos a la enseñanza en ingeniería. Del análisis, se eligieron las definiciones, propiedades y relaciones del intervalo de confianza con otros objetos matemáticos, así como representaciones, problemas y procedimientos de construcción de intervalos de confianza que serían objeto de evaluación. El estudio de los libros de texto se ha recogido en Olivo (2006) y una descripción de las pruebas piloto de algunos de los ítems del cuestionario, en Olivo y Batanero (2007).

Seguidamente se procedió a la elaboración y depuración de ítems que pudieran evaluar cada uno de los contenidos fijados. También se adaptó su formato y redacción para seguir las pautas marcadas en Osterlind (1989) sobre claridad del enunciado, número de distractores y formulación de la pregunta, y se realizaron pruebas de legibilidad con un grupo reducido de estudiantes. La colaboración de 10 expertos (profesores de estadística e investigadores en educación estadística) sirvió para seleccionar los ítems con los que finalmente se estructuraría el cuestionario, siguiendo el procedimiento de juicio de expertos. El cuestionario está compuesto por 12 ítems de opciones múltiples y seis problemas abiertos que, en su mayoría, fueron recopilados de diversas investigaciones.

En la muestra participaron un total de 252 estudiantes del tercer semestre de los estudios de ingeniería en el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM). Estos estudiantes estaban siguiendo un curso de probabilidad y estadística y habían estudiado los intervalos de confianza durante dos semanas. Asimismo, conocían las distribuciones de probabilidad, distribuciones muestrales y el resto de los temas incluidos en el cuestionario, los cuales les habían sido enseñados con anterioridad. Eran estudiantes de buen nivel, pues su calificación media en la asignatura el semestre anterior fue de 84 (sobre 100). Presentamos a continuación los resultados en los 12 ítems de opción múltiple y dos de los problemas abiertos que componen el cuestionario. Los resultados completos se recogen en Olivo (2008).

## ÍTEMES DE OPCIONES MÚLTIPLES

En el Anexo 1 se reproducen los ítems de opciones múltiples. El ítem 1 (Cruise, Dudley y Thayer, 1984) corresponde a la definición de intervalo de confianza, tra-

tando de detectar los siguientes sesgos: el intervalo se refiere a la media muestral (distractor *a*), relación incorrecta entre ancho del intervalo y coeficiente de confianza (distractores *b* y *d*). El ítem 2 (Gardfield, delMas y Chance, 2004) evalúa la comprensión del efecto del tamaño de muestra sobre la precisión (ancho de intervalos) cuando se mantiene constante el coeficiente de confianza. Los ítems 3 y 4 (Cruise, Dudley y Thayer, 1984) se refieren al efecto del coeficiente de confianza sobre el ancho del intervalo y su variación en diferentes muestras.

El ítem 5, de elaboración propia, requiere la capacidad de construcción de intervalos para la media de una población en una muestra grande con desviación típica  $\sigma$  conocida. Para ello, los estudiantes debían sumar y restar a la media muestral el error estándar (la desviación típica poblacional dividida por la raíz cuadrada del tamaño de muestra) multiplicado por el valor crítico de la distribución normal estándar. La respuesta correcta es la *b*. Los distractores comprueban el olvido del valor crítico de la distribución normal estándar que multiplica al error estándar (*a*) o la raíz cuadrada de la muestra que divide a la desviación típica poblacional (*c*) o ambas cosas (*d*). El ítem 6 tomado de Miller, Freund y Johnson (1997, p. 223) evalúa la capacidad para estimar la media de una población a partir de datos experimentales  $\sigma$  desconocida, muestra grande. La opción correcta es la *b*. Las opciones restantes sirven para detectar los siguientes conflictos: uso incorrecto de la distribución normal estándar para obtener el punto crítico (*a*); no dividir la desviación estándar muestral entre *n* (*c*) o ambos errores (*d*).

El ítem 7 se ha tomado de Devore (2005) y pide el cálculo del intervalo para la diferencia *de medias en dos poblaciones*, conociendo sus varianzas  $\sigma_1^2$  y  $\sigma_2^2$  y siendo las muestras independientes. La opción correcta *b* requiere que el alumno escriba la ecuación correcta del intervalo de confianza y aplique la distribución normal estándar para obtener los puntos críticos, utilizando el nivel de confianza correspondiente. Las opciones restantes detectan los siguientes errores: usar el valor crítico 2.57 en vez de 1.96 (opción *a*); no dividir entre *n* y *m* (opción *c*) y ambos errores (opción *d*).

El ítem 8, también adaptado de Devore (1998, 8*b*, p. 342), se refiere a la comparación de las medias en dos poblaciones,  $\sigma_1^2 = \sigma_2^2$  desconocidas, para el caso de muestras independientes grandes. La opción correcta es la *a*, donde se da la estimación correcta del intervalo y la interpretación correcta (tanto por ciento de muestras en la misma población e igual tamaño, cuyo intervalo calculado cubre el parámetro). Las opciones restantes corresponden a los siguientes errores: estimación correcta con interpretación incorrecta (opción *b*); interpretación correcta con estimación incorrecta (opción *c*); ambos errores (opción *d*).

El ítem 9 se ha adaptado de Montgomery y Runger (2004, p. 416) y evalúa la *comparación de dos varianzas poblacionales*. Al variar el intervalo entre valores menores que 1 y mayores que 1, cualquiera de los dos procesos se puede recomendar. La opción correcta es la *c*, que obtiene correctamente el intervalo de confianza del cociente de varianzas poblacionales y además da la interpretación correcta del intervalo. Las opciones restantes evalúan los siguientes errores: se obtiene una estimación correcta del intervalo, pero proporciona una interpretación incorrecta del intervalo (opción *a*); se cambian los grados de libertad en el numerador y denominador, obteniéndose valores críticos incorrectos de la distribución *F*. También se da una interpretación incorrecta del intervalo, aunque ese intervalo hubiese resultado correcto (opciones *b* y *d*).

El ítem 10 es de elaboración propia y el contenido evaluado es elegir un modelo de distribución muestral del estadístico. La opción correcta es la *b* y las restantes evalúan los siguientes errores: elegir una distribución que se aplica para la estimación de la media también en condiciones de muestras pequeñas (opción *a*); utilizar una distribución que se utiliza en la construcción de intervalos de confianza para la varianza sólo para muestras grandes (opción *c*), utilizar una distribución que solamente se aplica para la construcción de intervalos de confianza para cociente de varianzas poblacionales (opción *d*).

El ítem 11 es una adaptación de un problema de Cruise, Dudley y Thayer (1984). Se trata de determinar valores críticos en la distribución del estadístico. La opción correcta es la *c*, que obtiene los puntos críticos del intervalo de confianza para la media a partir de la distribución *t*, para esas condiciones particulares. Las opciones restantes evalúan los siguientes errores: suponer una distribución que se aplica sólo en condiciones de muestras grandes y, además, error en el uso del valor del nivel de confianza (opción *a*); suponer una distribución que se aplica en condiciones de muestras grandes (opción *b*); se usa la distribución correcta, pero se utiliza incorrectamente el valor de los grados de libertad de la distribución (opción *d*).

El último ítem, tomado de Behar (2001), evalúa la interpretación de gráficos de intervalos de confianza. La opción correcta es la *d*, puesto que, al producirse solapes de los intervalos, las medias de las poblaciones no difieren. Las opciones restantes estudian los siguientes errores: dar la interpretación contraria (opción *a*); interpretación inadecuada de la variabilidad para esos dos intervalos (opción *b*).

En el cuadro 1 presentamos la frecuencia de respuestas a cada opción de cada ítem, la proporción de respuestas correctas y el índice de discriminación, definido como diferencia en proporción de aciertos entre el grupo que tiene puntuación

**Cuadro 1** Resultados\* en los ítems de opción múltiple ( $n = 252$ )

Ítem	Frecuencia				% correcto	Índice discriminación	Contenido del ítem
	A	B	C	D			
1	65	40	<b>140</b>	7	55.6	0.13	Definición
2	84	<b>139</b>	28	1	55.2	0.30	Variación con tamaño de muestra
3	7	66	<b>164</b>	15	65.1	0.30	Relación con coeficiente de confianza
4	73	86	<b>92</b>	-	36.5	0.29	Variación en diferentes muestras
5	7	<b>199</b>	11	21	79.0	0.42	Estimar media $\sigma$ conocida, población normal
6	29	<b>184</b>	15	11	73.0	0.40	Estimar media $\sigma$ desconocida, muestra grande
7	18	<b>188</b>	19	9	74.6	0.34	Comparar medias, varianzas conocidas
8	<b>74</b>	81	48	42	29.4	0.33	Comparar medias, varianza desconocida, muestra grande
9	35	47	<b>133</b>	31	52.8	0.42	Comparar varianzas, población normal
10	78	<b>133</b>	25	15	52.8	0.39	Elegir distribución muestral
11	12	104	<b>118</b>	13	46.8	0.32	Determinar valor crítico
12	23	17	32	<b>175</b>	69.4	0.35	Interpretar gráficos intervalos

\* Soluciones correctas en negrita.

superior e inferior en el total de la prueba (dividido el grupo en tres partes de acuerdo con su puntuación). Se considera discriminativo el ítem si la diferencia es mayor que +0.3.

Como resumen del análisis de los ítems de opciones múltiples, podemos observar que los ítems fueron, en general, discriminativos, a excepción del primero, que recoge la definición del intervalo de confianza. Los ítems fueron bien respondidos por los estudiantes, los más fáciles fueron los ítems 5, 6 y 7, que evalúan los siguientes contenidos: a) *estimar la media de una población normal o en una muestra grande con  $\sigma$  conocida*, b) *comparar las medias en dos poblaciones, conociendo  $\sigma_1^2$  y  $\sigma_2^2$ , muestras independientes*, y c) *estimar la media de una población a partir de datos experimentales  $\sigma$  desconocida, muestra grande*, respectivamente.

Los más difíciles fueron los ítems 8, 4 y 11 que evalúan los siguientes contenidos: *a) comparar las medias en dos poblaciones  $\sigma_1^2 = \sigma_2^2$  desconocidas, muestras independientes grandes, b) significado del nivel de confianza (variación del intervalo en diferentes muestras), y c) determinar valores críticos en la distribución del estadístico.* Respecto a los principales errores detectados, los clasificamos a continuación de acuerdo con las entidades primarias consideradas en nuestro marco teórico:

#### *Comprensión de definiciones*

- Un 25.8% de estudiantes confunde estadístico y parámetro, pensando que el intervalo se construye para estimar la media muestral (ítem 1). Vallecillos y Batanero (1997), en su estudio de la comprensión de los contrastes de hipótesis estadísticas, informan también esta confusión. Mencionan que no se trata de un problema de uso incorrecto de la notación adecuada para cada concepto, sino al hecho de no tomar en consideración las distintas medias y distribuciones implicadas, en concreto, la distribución muestral del estadístico. Este error también ha sido encontrado por Behar (2001) y delMas, Garfield, Ooms y Chance (2007).

#### *Comprensión de propiedades*

- Un 33.3% de los estudiantes no comprenden cómo varía el ancho del intervalo al disminuir el tamaño de la muestra (ítem 2) y 26.2% de los estudiantes no comprenden como varía el ancho del intervalo al reducirse el nivel de confianza (ítem 3). Estos resultados confirman los obtenidos por Fidler y Cumming (2005) y Behar (2001). Sólo 36.5% comprende que el intervalo de confianza representa el porcentaje de intervalos de muestras, tomadas todas bajo las mismas condiciones, dentro de los cuales estará contenido el verdadero valor del parámetro (ítem 4), resultado que corrobora los obtenidos por Behar (2001).
- El 32.1% de los estudiantes obtiene correctamente el intervalo de confianza para diferencia de medias poblacionales, pero incurre en error al hacer la interpretación bayesiana (ítem 8). Aunque Behar (2001) también estudia la interpretación de intervalos de confianza para diferencia de medias, el ítem que él evalúa plantea como proposición correcta una diseñada bajo el enfoque de interpretación bayesiana: “Nosotros no sabemos el verdadero aumento medio en la producción, pero estamos 95% seguros de que el aumento medio en la producción ha quedado atrapado por este intervalo”

(Behar, 2001, p. 223). De ahí que el resultado de nuestro estudio sea un nuevo aporte a la investigación empírica de las dificultades en la comprensión de los intervalos de confianza.

#### *Comprensión de procedimientos*

- Un 31% de los estudiantes escogieron incorrectamente la distribución  $t$  de Student en la construcción de intervalos de confianza para la varianza en muestras pequeñas (ítem 10), siendo la distribución chi cuadrada la apropiada para esas condiciones.
- Un porcentaje alto (41.3%) de los estudiantes determinó un valor crítico en forma incorrecta al hacerlo a partir de la distribución normal estándar, cuando la distribución requerida era la distribución  $t$  de Student (ítem 11). Las dificultades encontradas a través de las respuestas a los ítems 10 y 11 han sido estudiadas en un contexto más general por Schuyten (1991), quien señala que la utilización simultánea de conceptos con diferentes niveles de concreción supone una dificultad para los estudiantes. Por ejemplo, la comprensión de la media, la media de la muestra y la de la población tienen diferentes niveles de abstracción que confunden y complican el trabajo de los estudiantes. Nuestros resultados confirman los señalamientos de Schuyten (1991).
- El 18.7% de los estudiantes obtiene incorrectamente el intervalo de confianza para la comparación de dos varianzas poblacionales, porque intercambia indebidamente los grados de libertad del numerador con los del denominador en sus lecturas de las tablas para obtener los valores críticos en la distribución  $F$  (ítem 9). Una posible explicación de este error es que no llegan a interiorizar el significado de los valores críticos. No hemos encontrado este error en otros trabajos, por lo cual este resultado también es una nueva aportación a la investigación en el tema.

#### *Comprensión de representaciones*

- El 12.7% de los estudiantes interpreta incorrectamente los intervalos de confianza a partir de un gráfico, mostrando poca capacidad de lectura de gráficos, ya que se elige con frecuencia el distractor  $c$  (ítem 12). Behar (2001), en su formato de falso o verdadero, no incluye en su conjunto de ítems alguno parecido al distractor  $c$ , aunque sí los otros distractores.
- Otros errores, que son a la vez procedimentales y se relacionan con las representaciones que tienen que ver con la obtención incorrecta del inter-

valo de confianza, son que los estudiantes no dividen las varianzas  $\sigma_1^2$ ,  $\sigma_2^2$ , entre los tamaños de muestra en la construcción del intervalo para la diferencia de medias poblacionales conocidas, muestras independientes grandes (7.5%, ítem 7) y olvidan en la fórmula del intervalo de confianza, para la media poblacional,  $\sigma$  conocida, el valor crítico de la distribución normal estándar (8.3%, ítem 5). Estos dos tipos de errores, que no hemos encontrado registrados en otras investigaciones, los podríamos explicar por falta de conectividad entre los símbolos y los conceptos a los que se refieren.

## PROBLEMAS ABIERTOS

El cuestionario también contenía algunos ítems abiertos, con objeto de evaluar las estrategias de resolución y argumentación de los estudiantes. A continuación analizamos dos de estos problemas y las soluciones aportadas por los estudiantes.

**Problema 1.** Un fabricante asegura que sus garrafones contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones, se determinó que, en promedio, contenían 0.94 litros de cloro puro, con desviación estándar de la muestra de 0.097. Construir un intervalo de confianza de 95% para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población. (La distribución del contenido de cloro por botella puede considerarse normal.)

Este *problema*, tomado de Cruise, Dudley y Thayer (1984), evalúa el conocimiento procedimental en la construcción de intervalos de confianza para la media de una población normal con desviación típica no conocida. Se han clasificado las soluciones encontradas según el siguiente criterio:

*Solución correcta:* puesto que el tamaño de la muestra es pequeño y  $\sigma$  es desconocida, se debe usar la distribución  $t$  de Student. En el ejemplo que presentamos a continuación, el alumno recuerda la distribución muestral de la media para el caso de muestras pequeñas; al referirse al subíndice  $\frac{\alpha}{2}$  manifiesta que comprende la relación del valor crítico con el nivel de confianza 95%. Recuerda las probabilidades necesarias para calcular el intervalo, sustituye los datos del problema

y calcula los valores críticos en la distribución de  $t$  de Student, usando correctamente las tablas. Por último, lleva a cabo varias operaciones algebraicas con inecuaciones para llegar finalmente al intervalo pedido.

Como no se conoce la desviación típica de la población, para calcular el intervalo de confianza tenemos que usar la distribución  $t$ . Los grados de libertad son  $n - 1 = 15$ . La desviación estándar de la muestra es  $S = 0.097$ . El intervalo de confianza viene dado por la siguiente fórmula:  $\bar{x} \pm t_{\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}$ , donde  $t_{\frac{\alpha}{2}}$  es el valor crítico correspondiente al coeficiente de confianza  $1 - \alpha$ ,  $s$  la desviación típica de la muestra y  $n$  el tamaño de la muestra. El valor de la media muestral es 0.94 y el tamaño de la muestra es 16, por lo que, buscando en tablas en 15 grados de libertad y  $\frac{\alpha}{2} = 0.025$ , obtengo el valor crítico de  $t$ , con lo cual tengo todos los datos para sustituirlos en la fórmula:

$$0.94 - (2.13) \frac{0.097}{\sqrt{16}} < \mu < 0.94 + (2.13) \frac{0.097}{\sqrt{16}}$$

realizando las operaciones, se llega al intervalo  $0.8884 < \mu < 0.9916$  (Alumno BM).

*Respuesta correcta con errores procedimentales.* Además de esta solución correcta, hemos encontrado una variante en la que el alumno escribe correctamente la fórmula. Al realizar el proceso, obtiene correctamente el valor crítico de  $t_{\frac{\alpha}{2}}$ , pero al completar las operaciones algebraicas, tiene un error de proce-

dimiento en el manejo de inecuaciones en el último paso; entonces el intervalo obtenido es incorrecto. Se reproduce un ejemplo:

$$n = 16, s = 0.097, 1 - \alpha = 0.95, \bar{x} = 0.94$$

$$\bar{x} \pm t_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} = 0.94 \pm (2.131) \frac{0.097}{\sqrt{16}} = 0.94 \pm 0.0516$$

$$-0.0515 \leq \mu \leq 0.9916 \text{ (Alumno MR).}$$

*Confusión conceptual en el concepto grados de libertad.* El estudiante plantea el intervalo con la distribución muestral correcta, pero aparece un error de comprensión del concepto “grados de libertad” de la distribución  $t$ , aunque realiza el resto del procedimiento correctamente, como se muestra en el caso siguiente:

$$\bar{x} \pm t \times \frac{s}{\sqrt{n}}; 0.94 \pm (2.11) \frac{0.097}{\sqrt{16}}; 0.089 < \mu < 0.991 \text{ (Alumno HE).}$$

*Error procedimental en el cálculo de valores críticos.* El alumno usa los grados de libertad correctos, pero tiene un error en el cálculo de valores críticos de  $t_{\frac{\alpha}{2}}$ , al tomar 5% a cada lado de las colas de la distribución  $t$  de Student en vez de 2.5%, como se aprecia en el ejemplo reproducido a continuación:

$$n = 16, \bar{x} = 0.94, s = 0.097, 0.94 \pm 1.75 \frac{(0.097)}{\sqrt{16}}, 0.8975 < \mu < 0.9824 \text{ (Alumno LO).}$$

*Confusión conceptual entre media poblacional y muestral.* El estudiante escribe en el intervalo la media muestral en lugar de la media poblacional. Además, puede cometer otros errores; en el ejemplo, el estudiante realiza cálculo incorrecto de valores críticos de  $t_{\frac{\alpha}{2}}$ , al tomar 5% a cada lado de las colas de la distribución  $t$  en lugar de 2.5%.  $\frac{\alpha}{2}$

$$n = 16, s = 0.097, \bar{x} = 0.94, 0.94 \pm 1.75 \frac{(0.097)}{\sqrt{16}}, 0.8975 < \bar{x} < 0.9824 \text{ (Alumno AC).}$$

*Confunde la distribución muestral.* El alumno escribe la fórmula del intervalo de confianza para la media, utilizando el valor crítico de  $z_{\frac{\alpha}{2}}$  en vez de  $t_{\frac{\alpha}{2}}$ .

En el ejemplo que se reproduce, el estudiante usa la distribución normal para obtener los valores críticos, en vez de la distribución  $t$ . Se presenta un conflicto conceptual al confundir la desviación estándar de la muestra con la desviación estándar de la población. Tanto en este caso como en el anterior, se produce una confusión entre ejemplar (media) y tipo (media de la muestra, media de la población); error que ha sido descrito, entre otros, por Vallecillos (1994).

$$n = 16, s = 0.097, \bar{x} = 0.94, 1 - \alpha = 0.95, \frac{\alpha}{2} = 0.025,$$

$$\bar{x} \pm \frac{\left(\frac{z_{\alpha}}{2}\right)\sigma}{\sqrt{n}}, \quad 0.94 \pm \frac{\left(\frac{z_{\alpha}}{2}\right)0.097}{\sqrt{16}}, \quad 0.94 \pm \frac{(1.96)0.097}{\sqrt{16}}$$

$$0.94 \pm 0.047, (0.893, 0.987) \text{ (Alumno LK).}$$

Otros errores consisten en escribir en su solución la fórmula del intervalo de confianza para la varianza, escribir de manera incompleta el intervalo o conflicto relativo al procedimiento de cálculo al sustituir en la posición del valor crítico de  $z_{\frac{\alpha}{2}}$  el valor de probabilidad de 95%. En el cuadro 2 presentamos los resultados

del problema 1 que resolvieron correctamente un 39.6%. Una tercera parte de los estudiantes contestó de manera parcialmente correcta, por lo cual lo consideramos un problema de dificultad moderada para los estudiantes.

El 34.5% de los estudiantes tuvo conflicto en la obtención de los grados de libertad de la distribución  $t$ , tomando  $n$  en lugar de  $n - 1$ , lo que muestra que el concepto de grados de libertad no se comprende. Hoy día, esta confusión se resuelve porque los programas estadísticos realizan el cálculo automático y no hay que especificar los grados de libertad. Un 19.5% de estudiantes resolvió el problema incorrectamente. El 3.2% tuvo errores en la obtención de los valores críticos a partir de las tablas, al tomar 5% a cada lado de las colas de la distribución  $t$  en vez de 2.5%, lo que es una dificultad no vinculada al concepto de intervalo de confianza, sino a la falta de comprensión de la idea de valor crítico.

Un 1.2% escribe en el intervalo media muestral, en vez de media poblacional además de errores procedimentales en la obtención de los valores críticos de  $t_{\frac{\alpha}{2}}$ , al tomar 5% a cada lado de las colas de la distribución  $t$  en vez de 2.5%. Un

11.5% confunde la desviación estándar de la muestra con la desviación estándar poblacional, por lo que utilizan la distribución normal para calcular los valores críticos en vez de utilizar la distribución  $t$ . Un 2.8% (dentro de la clasificación de otros errores) comete error al escribir la fórmula del intervalo de confianza para la varianza, pero el parámetro que escribe al centro del intervalo es el de la media poblacional, con lo cual el alumno relaciona inadecuadamente los parámetros media y varianza con sus límites de confianza correspondientes. Finalmente, 0.8%

**Cuadro 2** Frecuencias (y porcentajes) de respuestas en el problema 1 ( $n = 252$ )

	Frecuencia	Porcentaje
Respuestas correctas	89	35.3
Respuestas correctas con errores procedimentales en inequaciones	11	4.4
Confusión conceptual en el concepto grados de libertad	87	34.5
Errores procedimentales en el cálculo de valores críticos	8	3.2
Confusión conceptual entre media poblacional y muestral	3	1.2
Confunde la distribución muestral	29	11.5
Otros errores	9	3.6
Blanco	16	6.3
Total	252	100.0

(dos alumnos) tienen un error en el procedimiento de cálculo del valor crítico, escriben el valor de probabilidad de 0.95 en lugar del valor crítico de  $t$ .

**Problema 2.** Sea  $\sigma^2$  la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es  $s^2 = 13\,700$ ,  $n = 16$ . Calcular el intervalo de confianza de 95% para  $\sigma$ .

Este problema, tomado de Devore (2005, p. 310), evalúa el conocimiento del procedimiento de construcción de intervalos para la estimación de una varianza. Las respuestas obtenidas son las siguientes:

*Respuesta correcta.* Puesto que el tamaño de la muestra es pequeño, se debe usar la distribución Chi-cuadrado y luego seguir pasos análogos a los descritos en el problema 1. El razonamiento correcto que esperamos es el siguiente:

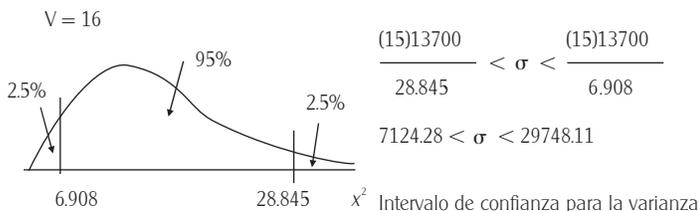
Sea  $\sigma^2$  la varianza. Sabemos que se verifica 
$$\frac{(n-1)S^2}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2},$$

entonces 
$$\frac{(16-1)13700}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(16-1)13700}{\chi_{1-\alpha/2}^2},$$
 es decir, 
$$\frac{(16-1)13700}{27.488}$$

$$\sigma^2 < \frac{(16-1)13700}{6.262}$$

sustituyendo en las inecuaciones se llega al intervalo  $7\,975.9895 < \sigma^2 < 32\,816.991$ , luego obteniendo raíz cuadrada a ambos lados de la desigualdad, se obtiene el intervalo de confianza para la desviación estándar poblacional,  $89.308 < \sigma < 181.154$  (*Alumno AC*).

*Confusión conceptual en el concepto grados de libertad.* El alumno inicia correctamente la solución, pero los grados de libertad los obtiene sin restar 1 al valor del tamaño de la muestra, como en el caso siguiente. En este ejemplo, el alumno usa una representación gráfica de la distribución adecuada para este intervalo, dividiendo además la probabilidad total en tres partes, pero al avanzar surge un problema relacionado con los procedimientos al calcular incorrectamente los grados de libertad de la distribución Chi-cuadrado. Además hay una confusión de la notación: (*Alumno EG*)



*Error procedimental al olvidar la expresión de los denominadores.* El estudiante

escribe el intervalo  $\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}$ , olvidando las expresiones en

los denominadores del intervalo, luego continúa en forma correcta el resto del procedimiento. Como consecuencia, obtiene un intervalo incorrecto para la desviación estándar poblacional. Podemos observarlo en el siguiente ejemplo.

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}, \quad \frac{(16-1)13700}{\chi^2_{\left(1-\frac{\alpha}{2}\right)}} < \sigma^2 < \frac{(16-1)13700}{\chi^2_{\frac{\alpha}{2}}}$$

(*Alumno DL*).

*Error procedimental:* el alumno recuerda solamente un límite de la fórmula del intervalo de confianza para la varianza. Sustituye los datos y realiza algunas operaciones, pero sólo obtiene uno de los límites. En el ejemplo que presentamos, también hay un error en la obtención de los grados de libertad de la distribución Chi-cuadrado.

$$\sigma^2 = \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2},v}^2} = \frac{(15)(13700)}{\chi_{1-\frac{\alpha}{2},v}^2}, (1-\alpha)\% = 95\%; \alpha = 1 - 95\%; \alpha = 5\%$$

$$\frac{15(13700)}{\chi_{0.025,16}^2} = \frac{205500}{28.845} = 7124.284 \text{ (Alumno MV).}$$

*Error procedimental al confundir la expresión del intervalo.* En este grupo están las respuestas de los alumnos que escriben una fórmula del intervalo de confianza para la varianza que se parece más bien a la fórmula del intervalo para la media poblacional. También se incluyen las respuestas de los que escriben una fórmula que se acerca a la de la varianza, pero en donde confunden la distribución, escribiendo  $\frac{z_{\frac{\alpha}{2}}}{2}$  en vez de escribir  $S^2$ , escriben

$$\frac{(n-1)z_{\frac{\alpha}{2}}}{\chi_{\frac{\alpha}{2}}^2} < \sigma^2 < \frac{(n-1)z_{\frac{\alpha}{2}}}{\chi_{\left(1-\frac{\alpha}{2}\right)}^2}.$$

En el cuadro 3 presentamos los resultados del problema 2 (calcular un intervalo de confianza para la desviación estándar poblacional). El problema resultó difícil para los estudiantes, pues solamente 38.5% respondieron de manera correcta. Un 16.7% tienen un error conceptual, ya que calculan incorrectamente los grados de libertad de la distribución Chi-cuadrado. Podríamos suponer que el exceso de variedad de símbolos utilizados en la obtención de intervalos de confianza provoca que el alumno se pierda y algo tan sencillo como es el concepto de grados de libertad de la distribución no alcanza a ser retenido por los estudiantes. Para esta proporción de estudiantes, ese problema se magnifica con otro error, al escribir desviación estándar en vez de varianza. Sea desatención o no, lo que se muestra es que los estudiantes no asocian los símbolos a los objetos matemáticos que intervienen en estos procesos.

**Cuadro 3** Frecuencias (y porcentajes) de respuestas en el problema 2 ( $n = 252$ )

	Frecuencia	Porcentaje
Respuestas correctas	97	38.5
Confusión conceptual en el concepto de grados de libertad	42	16.7
Error procedimental: olvido del denominador	15	6.0
Error procedimental: sólo recuerda un límite	70	27.8
Error procedimental: confunden la expresión del intervalo	10	4.0
Blanco	18	7.1
Total	252	100.0

Aparece una proporción mayor de errores procedimentales: un 6% de los alumnos tiene un conflicto procedimental al olvidar parte de la fórmula del intervalo de confianza para la varianza. Un 31.8% comete otros errores de procedimiento, destacan los alumnos que recuerdan solamente un límite del intervalo (27.8%). La familiaridad que alcanzaron con las fórmulas del intervalo de confianza para la media poblacional no logran transferirla a la varianza.

## DISCUSIÓN Y CONCLUSIONES

De acuerdo con nuestra perspectiva teórica, los resultados del estudio sugieren la complejidad de la comprensión del objeto “intervalo de confianza”, donde aparecen una serie de dificultades y errores de comprensión de las entidades primarias que lo componen.

Respecto a la *comprensión de definiciones*, confirmamos en el ítem 1 y en el problema 1 la confusión entre los conceptos de estadístico y parámetro descrita por Vallecillos y Batanero (1997). Interpretamos este error en términos de confusión entre ejemplar (media de la muestra, media de la población) y tipo (media), los cuales no se diferencian.

Tampoco se comprende la definición de grados de libertad, lo que causa errores en la construcción de los intervalos de confianza en los problemas 1 y 2.

Respecto a la *comprensión de propiedades*, una tercera parte de los alumnos mostraron conflictos de comprensión de cómo se relaciona el tamaño de la muestra con el ancho del intervalo (ítem 2), resultado que nos revela un aspecto

estructural de los significados personales de los estudiantes (Budé, 2006) y corrobora los obtenidos por Fidler y Cumming (2005) y Behar (2001). Tampoco comprenden que los intervalos de confianza representan el porcentaje de intervalos de muestras, tomadas todas en las mismas condiciones, dentro de los cuales estará contenido el verdadero valor del parámetro (ítem 4), con lo que se confirman las investigaciones de Behar (2001).

Una aportación de nuestro trabajo es comprobar que una proporción importante de estudiantes da una interpretación bayesiana a dicha definición, suponiendo que los extremos del intervalo son fijos y no aleatorios (ítem 8). Aunque Vallecillos y Batanero (1997) han descrito este error en relación con el contraste de hipótesis, nuestro trabajo aporta la detección de este conflicto para el caso del intervalo de confianza.

En relación con la *comprensión procedimental*, una parte de los estudiantes no demostraron habilidad para seleccionar una distribución de muestreo apropiada (ítem 10 y problema 1) o hacen un uso inadecuado de las tablas para la determinación de valores críticos (ítem 11 y problema 1). En el ítem 9 se intercambian los grados de libertad de numerador y denominador en la distribución  $F$ . Ninguno de estos errores ha sido descrito en las investigaciones previas. Hemos encontrado también otros errores procedimentales de menor importancia, como el olvido de los grados de libertad en las distribuciones  $t$  y Chi-cuadrado o de algunos de los factores en las fórmulas de cálculo de los intervalos de confianza, tanto en los ítems como en los problemas abiertos. En dichos problemas, otros fallos procedimentales se deben al manejo inadecuado de inecuaciones, olvido de las fórmulas del intervalo o escritura incompleta del intervalo en el problema 2.

Respecto a la *comprensión de representaciones*, algunos alumnos tienen dificultades al interpretar gráficamente los intervalos de confianza en el ítem 12, tomado de Behar (2001). Hemos encontrado también errores procedimentales causados por falta de conectividad entre los símbolos y los conceptos a los que se refieren.

En resumen, nuestro estudio de evaluación confirma los resultados obtenidos en otros trabajos en relación con la comprensión de las distribuciones muestrales (Schuyten, 1991; Vallecillos y Batanero, 1997; delMas, Garfield, Ooms y Chance, 2007). Al mismo tiempo, los amplía y detecta otras dificultades de comprensión, como la interpretación bayesiana del significado del coeficiente de confianza, las dificultades con la obtención de valores críticos y de interpretación de salidas de computadora.

También ponemos de manifiesto, mediante el análisis realizado de las soluciones correctas e incorrectas a los problemas abiertos, la complejidad de la

construcción del intervalo de confianza y la multitud de objetos matemáticos que el estudiante ha de poner en relación. En nuestra búsqueda de investigaciones relacionadas, encontramos un trabajo de Newcombe (1998) que estudia la construcción del intervalo de confianza para una proporción, pero no explora las dificultades de comprensión en los estudiantes. No hemos encontrado estudios relacionados con la construcción de intervalos para la varianza, por lo cual, nuestro resultado es una aportación a la investigación de las dificultades en la comprensión y construcción de éstos.

La importancia del intervalo de confianza y las recientes sugerencias de la necesidad de su uso en la investigación recomiendan tener en cuenta estos resultados y tratar de mejorar su enseñanza. El análisis realizado permite mostrar los diferentes elementos que intervienen en su comprensión y que han de ser objeto de enseñanza diferenciada. Una posibilidad de mejorar esta enseñanza sería apoyarse en la simulación con computadoras, donde los estudiantes puedan explorar el significado de los intervalos y el efecto que tiene sobre ellos el tamaño de la muestra, la varianza y el coeficiente de confianza (Terán, 2006).

Asimismo, sería necesario hacer conscientes a los estudiantes de la interpretación correcta de dicho coeficiente y de la diferencia entre estadístico y parámetro. De igual manera, se debe relacionar más estrechamente este tema con el estudio de las distribuciones de probabilidad y hacer ver a los estudiantes la importancia de elegir una adecuada distribución muestral en el cálculo de los intervalos de confianza.

## AGRADECIMIENTOS

Este trabajo forma parte del proyecto SEJ2007-60110 (MEC-Feder).

## REFERENCIAS BIBLIOGRÁFICAS

- Batanero, C. (2000), "Controversies around significance tests", *Journal of Mathematics Thinking and Learning*, vol. 2, núms. 1-2, pp. 75-98.
- Behar, R. (2001), *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística*, Tesis doctoral, Universidad Politécnica de Cataluña.
- Budé, L. (2006), "Assessing students' understanding of statistics", en A. Rossman y Beth Chance (eds.), *Proceedings of the Seventh International Conference*

- on *Teaching of Statistics*, CD-ROM, Salvador (Bahía), Brasil, International Association for Statistical Education.
- Cruise, R., R. Dudley y J. Thayer (1984), *A Resource Guide for Introductory Statistics*, Nueva York, Kendall/Hunt.
- Cumming, G. y S. Finch (2005), "Inference by eye: Confidence intervals, and how to read pictures of data", *American Psychologist*, núm. 60, pp. 170-180.
- Cumming, G., J. Williams y F. Fidler (2004), "Replication and researchers' understanding of confidence intervals and standard error bars", *Understanding Statistics*, núm. 3, pp. 299-311.
- delMas, R.C., J.B. Garfield, A. Ooms y B.L. Chance (2007), "Assessing students' conceptual understanding after a first course in statistics", *Statistics Education Research Journal*, vol. 6, núm. 2, pp. 28-58, [www.stat.auckland.ac.nz/serj](http://www.stat.auckland.ac.nz/serj).
- Devore, J. (1998), *Probabilidad y estadística para ingeniería y ciencias*, 4a. ed., México, Thomson.
- (2005), *Probabilidad y estadística para ingeniería y ciencias*, 6a. ed., México, Thomson.
- Fidler, F. y G. Cumming (2005), "Interval estimates for statistical communication: problems and possible solutions", trabajo presentado en la *IASE Satellite Conference on Communication of Statistics*, Sydney, International Association for Statistical Education.
- Garfield, J.B., R.C. delMas, y B.L. Chance (1999), "The role of assessment in research on teaching and learning statistics", trabajo presentado en el *American Educational Research Association. Annual Meeting*, Montreal, Canadá.
- (2004), *Tools for Teaching and Assessing Statistical Inference*, [http://www.gen.umn.edu/research/stat\\_tools](http://www.gen.umn.edu/research/stat_tools).
- Godino, J. D. (2002), "Un enfoque ontológico y semiótico de la cognición matemática", *Recherches en Didactique des Mathématiques*, vol. 22, núms. (2 y 3), pp. 237-284.
- Godino, J.D. y C. Batanero (1998), "Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education", en A. Sierpinska y J. Kilpatrick (eds.), *Mathematics Education a Research Domain: A Search for Identity*, Dordrecht, Kluwer, pp. 177-195.
- Harlow, L., S.A. Mulaik y J.H. Steiger (1997), *What if there Were no Significance Tests?*, Mahwah, NJ, Lawrence Erlbaum.
- Miller, I., J. Freund y R. Johnson (1997), *Probabilidad y estadística para ingenieros*, 5a. ed., México, Prentice Hall.

- Montgomery, D. y G. Runger (2004), *Probabilidad y estadística aplicadas a la ingeniería*, 2a. ed., México, Limusa.
- Morrison, D.E. y R.E. Henkel (1970), *The Significance Test Controversy*, Chicago, Aldine.
- Newcombe, R. G. (1998), "Two-sided confidence intervals for the single proportion: Comparison of seven methods", *Statistics in Medicine*, núm. 17, pp. 857-872.
- Olivo, E. (2006), *Análisis de la presentación de intervalos de confianza en textos de estadística para ingenieros*, trabajo de investigación tutelada, Universidad de Granada.
- (2008), *Significados de los intervalos de confianza para los estudiantes de ingeniería en México*, Tesis doctoral, Universidad de Granada.
- Olivo, E. y C. Batanero (2007), "Un estudio exploratorio de dificultades de comprensión del intervalo de confianza", *UNION*, núm. 12, [www.sinewton.org/numeros/](http://www.sinewton.org/numeros/).
- Osterlind, S. J. (1989), *Constructing Test Items*, Boston, Kluwer.
- Schenker, N. y J.F. Gentleman (2001), "On judging the significance of differences by examining the overlap between confidence intervals", *The American Statistician*, núm. 55, pp. 182-186.
- Schuyten, G. (1991), "Statistical thinking in psychology and education", en D. Vere-Jones (ed.), *Proceeding of the Third International Conference on Teaching Statistics*, Voorburg, Países Bajos, International Statistical Institute, pp. 486-490.
- Terán, T. (2006), "Elements of meaning and its role in the interaction with a computational program", en A. Rossman y B. Chance (eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador (Bahía), Brasil, IASE.
- Vallecillos, A. (1994), *Estudio teórico-experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*, Tesis doctoral, Departamento de Didáctica de la Matemática, Universidad de Granada.
- Vallecillos, A. y C. Batanero (1997), "Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitarios", *Recherches en Didactique des Mathématiques*, vol. 17, núm. 1, pp. 29-48.

## ANEXO 1. ÍTEMS DE OPCIONES MÚLTIPLES

**Ítem 1.** El intervalo de confianza de 50% para la media de una población  $\mu$  es:

- El rango dentro del cual cae 50% de los valores de la media de la muestra  $\bar{x}$ .
- Un intervalo más ancho que el intervalo de confianza de 95%.
- Un intervalo de valores calculado a partir de los datos de la muestra. En 50% de las muestras de una población, el intervalo calculado contiene a la media de la población.**
- Dos veces más ancho que el intervalo de confianza de 100%.

**Ítem 2.** Comparado con los intervalos de confianza calculados en muestras de tamaño  $n = 4$ , el ancho de los intervalos de confianza de la media de una población normal calculado en muestras de tamaño  $n = 50$ :

- Variará más que los anchos de los intervalos para muestras de tamaño  $n = 4$ .
- Variará, pero no tanto como los anchos de los intervalos para muestras de tamaño  $n = 4$ .**
- Tomarán valores parecidos.

**Ítem 3.** Si, manteniendo todos los demás datos fijos, el nivel de confianza se reduce (por ejemplo de 90% a 80%):

- El intervalo de confianza no cambia.
- El intervalo de confianza será más ancho.
- El intervalo de confianza será más angosto.**
- El cambio en el intervalo de confianza no es predecible.

**Ítem 4.** En un intervalo de confianza de 95% para la media:

- Si se toman muchas muestras y con cada una se construye el intervalo, la media muestral  $\bar{x}$  caerá dentro del intervalo de confianza 95% de las veces.
- La probabilidad de que  $\bar{x}$  caiga en un intervalo de confianza calculado de una muestra específica es 0.95.
- Si se toman muchas muestras de igual tamaño, 95% de los intervalos de confianza calculado contendrían a  $\mu$ .**

**Ítem 5.** La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza a 95% para la media de la población, asumiendo que  $\sigma = 7$ .

- a. (61.28, 88.72).
- b. (73.63, 76.37).**
- c. (68, 82).
- d. (74.3, 75.7).

**Ítem 6.** Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre para una muestra de tamaño  $n = 100$ , media:  $\bar{x} = 18$  y varianza muestral  $s^2 = 36$ . Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre.

- a. (17.016, 18.984).
- b. (16.824, 19.176).**
- c. (6.24, 29.76).
- d. (8.16, 27.84).

**Ítem 7.** El siguiente cuadro resume la información sobre la resistencia a la compresión de cubos ( $N/mm^2$ ) para especímenes de concreto. Suponga que las desviaciones estándar poblacionales de ambos grupos son  $\sigma_1 = 4.89$  y  $\sigma_2 = 6.43$ , respectivamente.

	Tamaño muestral	Media muestral
Tipo 1	68	26.99
Tipo 2	74	37.56

Calcule un intervalo de confianza de 95% para hallar la diferencia entre el verdadero promedio de resistencia en el tipo 1 y el verdadero promedio de resistencia en el tipo 2:

- a. (-13.02, -8.12).
- b. (-12.437, -8.70).**
- c. (-31.32, 10.18).
- d. (-18.64, -2.5).

**Ítem 8.** El cuadro resume un estudio de las características de tornillos de anclaje. Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte.

Diámetro de tornillo	Tamaño muestra	Resistencia al corte	
		Media muestral	Desviación estándar
3/8	100	4.25	1.3
1/2	100	7.25	1.7

- (−3.41, −2.58): en 95% de las muestras del mismo tamaño en la población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.**
- (−3.41, −2.58): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte caiga en el intervalo (−3.41, −2.58) es 0.95.
- (−3.35, −2.65): en 95% de las muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.
- (−3.35, −2.65): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte caiga en el intervalo (−3.35, −2.65) es 0.95.

**Ítem 9.** Una compañía quiere seleccionar el proceso de pulido que presente la menor variabilidad. Una muestra aleatoria de  $n_1 = 16$  piezas del primer proceso da una desviación estándar muestral  $s_1 = 5$  micropulgadas, y una muestra aleatoria de  $n_2 = 11$  piezas del segundo proceso da una desviación estándar muestral  $s_2 = 4$  micropulgadas. Establezca un intervalo de confianza de 90%

para  $\frac{\sigma_1^2}{\sigma_2^2}$ , suponiendo que los dos procesos son independientes y que la aspereza superficial tiene una distribución normal. ¿Cuál de los dos procesos recomendaría usted?

- Como todos los valores están en el intervalo (0.54, 3.96), recomendaría el proceso 1.
- Como todos los valores están en el intervalo (0.61, 4.44), recomendaría el proceso 2.

- c. **Como el cociente de valores está en el intervalo (0.54, 3.96), recomendaría cualquiera de los dos.**
- d. Como todos los valores están dentro del intervalo (0.61, 4.44), recomendaría cualquiera de los dos.

**Ítem 10.** La distribución muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas es:

- a. Distribución *t* de Student.
- b. **Distribución Chi-cuadrada.**
- c. Distribución normal.
- d. Distribución *F*.

**Ítem 11.** El nivel de confianza es 0.95, para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida para un grupo de puntajes distribuido normalmente de tamaño  $n = 20$ . Los valores críticos han de ser:

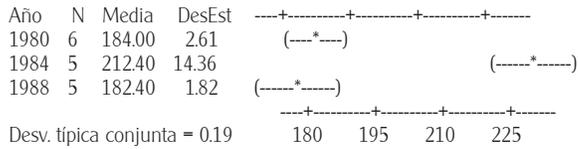
- a. -1.65 y 1.65.
- b. -1.96 y 1.96.
- c. **-2.093 y 2.093.**
- d. -2.085 y 2.08

**Ítem 12.** Considere el gráfico 1 del rendimiento medio de cebada en 1980, 1984 y 1988 junto con un intervalo de 95% de confianza respectivos. ¿Cuál de las siguientes afirmaciones es verdadera?:

- a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia de que las medias de las muestras difieran.
- b. La estimación de la media de la población en 1980 es menos precisa que en 1988.
- c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia de que las medias de las poblaciones respectivas difieran.
- d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia de que las medias difieran.**

**Gráfico 1**

Intervalos de 95% para la media



**DATOS DE LOS AUTORES**

**Eusebio Olivo**

Instituto Tecnológico y de Estudios Superiores de Monterrey, México  
 eusebio.olivo@itesm.mx

**Carmen Batanero**

Universidad de Granada, España  
 batanero@ugr.es

**Carmen Díaz**

Universidad de Huelva, España  
 carmen.diaz@dpsi.uhu.es