

Universidad Central de Venezuela
Facultad de Humanidades y Educación
Escuela de Educación
Centro de Investigaciones Educativas (CIES)



Investigación Educativa: **Venezuela en Latinoamérica Siglo XXI**

Parte II

Carlos E. Blanco (Compilador)

UNIVERSIDAD CENTRAL DE VENEZUELA
Facultad de Humanidades y Educación
Escuela de Educación
Centro de Investigaciones Educativas



Investigación Educativa:
Venezuela en Latinoamérica Siglo XXI
(Parte II)

CRÉDITOS

© **Copyright 2011 Centro de Investigaciones Educativas (CIES)**

Depósito Legal: lf 20520113704086

ISBN: 978-980-00-2694-6

Compilador: Carlos E. Blanco

Diseño y diagramación: José Luis Revete

Publicado por:

Centro de Investigaciones Educativas. Escuela de Educación, UCV

Edif. Tránsito, P.B. Telf. 605-3006 / 605 2953 Fax: 605-2952

<http://web.ucv.ve/cies>.

Correo electrónico: cies@ucv.ve, jec.informacion@gmail.com

AUTORIDADES DE LA
UNIVERSIDAD CENTRAL DE VENEZUELA

Cecilia García Arocha, Rectora
Nicolás Bianco, Vicerrector Académico
Bernardo Méndez, Vicerrector Administrativo
Amalio Belmonte, Secretario

FACULTAD DE HUMANIDADES Y EDUCACIÓN

Vicenzo Piero Lo Monaco, Decano
Mariángeles Páyer Sánchez, Coordinadora Académica
Eduardo Santoro, Coordinador Administrativo
Vidal Sáez, Coordinador de Postgrado
Marina Polo, Coordinadora de Investigación
Hugo Quintana, Coordinador de Extensión

ESCUELA DE EDUCACION

Nora Ovelar P., Directora
Eyra Valdivieso, Coordinadora Académica
Carmen Elena Chacón, Coordinadora Administrativa
Rosario Sánchez, Coordinadora de los Estudios Universitarios Supervisados
Marina Polo, Coordinadora de Componente de Formación Docente
Alejandra Fernández, Coordinadora de Postgrado de Educación
Carlos Manterola, Coordinador de Extensión
Audy Salcedo, Jefe del Centro de Investigaciones Educativas

13 Interpretación de Intervalos de Confianza por Docentes en Formación

Audy Salcedo
audy.salcedo@ucv.ve
Universidad Central de Venezuela

Amalio Sarco Lira B.
asarcolira@gmail.com
Universidad Central de Venezuela

Jesús González
jagonza05@gmail.com
Universidad Central de Venezuela

Gabriel Yáñez Canal
gyanez@uis.edu.co
Universidad Industrial de Santander

Recibido: 21/06/2011 Aprobado: 22/11/2011

Introducción

En este artículo se reporta una investigación sobre la interpretación que tienen docentes en formación a los intervalos de confianza y forma parte de una investigación más amplia sobre este tema. En los cursos de estadística inferencial en el ámbito universitario, el tema de los intervalos de confianza normalmente aparece después de trabajar las distribuciones muestrales y previo al contraste de hipótesis clásico. No obstante, es frecuente que los profesores otorguen mayor preponderancia al contraste de hipótesis, debido a su tradicional y profuso uso de esta técnica estadística en la investigación científica, así como su amplia aceptación como "criterio de validación" de los resultados de un experimento. Otros propulsores de este desequilibrio son los editores y árbitros de las revistas especializadas, quienes le dan preferencia a los artículos donde el contraste de hipótesis tenga como resultado $P < 0,05$, olvidando que Fisher vio el valor P como

un índice que mide la fuerza de la evidencia en contra de la hipótesis nula y no como un punto arbitrario de división entre resultados estadísticamente "significativos" y "no significativos". Se interpreta erróneamente la significación estadística asumiendo que necesariamente se halló un resultado importante y, por lo tanto, debe ser publicado.

Desde hace varios años diversos investigadores han llamado la atención sobre la inconveniencia de ese desequilibrio en el uso de dichas técnicas estadísticas y más recientemente han comenzado a sentirse en el ámbito educativo. Cohen (1994), Wilkinson (1999) y Thompson (2002) son algunos de los recientes críticos de lo que se conoce como el "ritual de la hipótesis nula" o la falacia del criterio de " P menor que 0,05". Estos investigadores consideran que centrar la atención en la decisión de si los resultados son o no estadísticamente significativos no es la única alternativa que tiene la investigación científica para presentar sus resultados. Ellos recomiendan no menospreciar la información que brindan los intervalos de confianza y el tamaño del efecto, consideran que esas medidas son el complemento adecuada a los resultados del contraste de hipótesis.

Cohen (1994), Wolfe y Cumming (2004) y Davies y Crombie (2009) abogan por el uso de los intervalos de confianza ya que con ellos se informa sobre el tamaño del efecto pero, además, proporcionan la misma información que el valor P . Por ejemplo, Wolfe y Cumming (2004) señalan que los intervalos de confianza son un medio eficaz de cuantificar la incertidumbre inherente a los resultados de un estudio, pero, además, tienen una estrecha relación con el valor P , presentar el intervalo de confianza del 95% o del 99% hace innecesario indicar si el correspondiente valor de P es inferior a 0,05 o a 0,01. También hay investigadores que creen importante presentar el tamaño del efecto, pero consideran inapropiado el uso de los intervalos de confianza ya que consideran que muchos de ellos no comprenden su significado (Killeen, 2006). No obstante, la recomendación de presentar el tamaño del efecto y los intervalos de confianza ha sido aceptada en asociaciones tan influyentes como la American Psychological Association (APA). Asimismo también ha comenzado a prestársele atención en la enseñanza, con la consecuente investigación didáctica sobre el tema. La importancia de este tipo de trabajo se encuentra en que pue-

den orientar a los docentes sobre las dificultades que presentan los estudiantes en la comprensión e interpretación de los intervalos de confianza y con ello una mejor selección de las actividades de aprendizaje que pueden proponer a sus estudiantes. Asimismo, la poca investigación desarrollada hasta los momentos sobre este tema, afirman Olivo y Batanero (2007), justifica la realización de investigaciones como la presente.

Intervalos de confianza

De manera general, un intervalo de confianza es un rango de valores aleatorios de una variable de interés, construido con una probabilidad previamente especificada de incluir el verdadero valor de la variable o parámetro de la población. La probabilidad especificada se conoce como nivel de confianza y los puntos extremos del intervalo se denominan límites de confianza. Al igual que el contraste de hipótesis clásico, los intervalos de confianza están fundamentados en el concepto de probabilidad frecuencial, lo cual implica regularidad cuando se considera un gran número de casos. Con el propósito de aclarar este planteamiento se presenta a continuación el caso de la construcción del intervalo de confianza para la media a partir de una muestra que proviene de una población normal con varianza conocida. Aunque se trata de un caso particular, es útil para ilustrar el razonamiento de construcción de intervalos de confianza y su interpretación.

Se estudia una población en la que se tiene una variable aleatoria con distribución $N(\mu, \sigma)$ con σ conocida. Se extrae una muestra de tamaño n y se desea estimar la media μ de la población por medio de un intervalo de confianza. El estimador puntual de μ es la media muestral cuya distribución muestral es conocida. La media de esta distribución es, precisamente, el parámetro μ y con una desviación estándar de $\frac{\sigma}{\sqrt{n}}$.

En este contexto de estimación se acostumbra llamar a esta desviación, error estándar. De aquí se deduce, por el proceso clásico de estandarización, que:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Para construir un intervalo de confianza, se utiliza la variable aleatoria Z , que depende tanto de la variable "media muestral" \bar{x} como del parámetro de interés μ . Y que tiene distribución normal estándar. Sobre esa distribución $N(0, 1)$ se selecciona el área $1-\alpha$ en el centro de la curva, de tal manera que el área restante en las dos colas es α , por lo tanto se tienen dos puntos equidistantes $-Z_{1-\frac{\alpha}{2}}$ y $Z_{1-\frac{\alpha}{2}}$, tales que

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq Z \leq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Sustituyendo Z por su valor en este caso particular, se obtiene

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Despejando la media muestral y la varianza, se obtiene

$$(A) \quad P\left(\bar{x} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Por lo tanto $\bar{x} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ como $\bar{x} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$,

que son los límites del intervalo, son cantidades aleatorias que dependen de la media muestral \bar{x} .

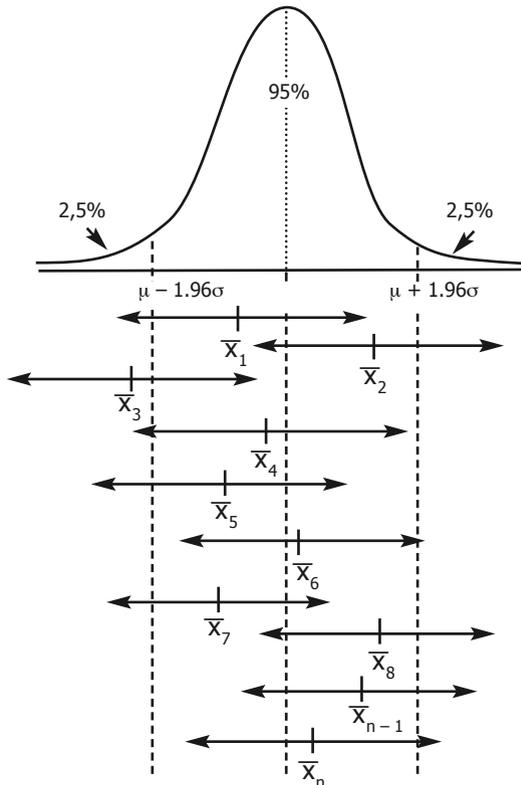
La expresión A permite construir el intervalo de confianza para un nivel de confianza $1 - \alpha$, que se expresa en términos de porcentaje como $100(1 - \alpha)\%$; contiene información sobre el estimador puntual de la media poblacional, es el valor central del intervalo y sobre el error de la estimación en términos del error estándar que depende directamente de la dispersión poblacional e inversamente proporcional al tamaño de la muestra. Por lo tanto:

- Al aumentar el nivel de confianza, manteniendo constantes los demás términos, se aumenta el tamaño del intervalo pero se disminuye la exactitud de la estimación.
- Al aumentar el tamaño muestral, dejando constantes los demás términos, se disminuye el tamaño del intervalo, por lo tanto se aumenta la exactitud de la estimación.
- Las poblaciones con menor dispersión permiten realizar estimaciones por intervalos más exactas.

En libros como los de Johnson y Kuby (2004) y Glass y Stanley (1994), luego explicar cómo calcular los intervalos de confianza es frecuente encontrar ejemplos de cálculo de intervalos de confianza, seguidos de ejercicios, todo esto con el propósito de desarrollar el conocimiento procedimental para la construcción de intervalos de confianza. Ese mismo esquema suelen seguirlo los profesores en sus clases, mientras que los estudiantes se preguntan: ¿cómo se puede confiar en la información de sólo una muestra? Para responderla el profesor debe dedicar un tiempo a recordar lo que significa la regularidad a largo plazo y volver a la distribución muestral de medias muestrales.

La distribución muestral de medias, se forma a partir de todas las muestras posibles de tamaño n a las cuales se le calcula la media. Si se seleccionan m muestras, se tendrán entonces m medias, es decir m valores, con los cuales se forma una distribución particular, la de las medias muestrales. Como se señaló antes esa distribución es normal, si la distribución de la variable en la población es normal independientemente del tamaño de la muestra, o bien si el tamaño de las muestras es mayor de treinta independientemente de la forma de la distribución original por el Teorema Central del Límite (TCL). Este resultado, permite asegurar, por ejemplo, que cerca del 95% de las medias muestrales, se encuentran dentro de dos desviaciones típicas a ambos lados de la media poblacional. Por lo tanto, si con cada una de las medias muestrales se construye un intervalo del 95% de confianza, el 95% de ellos contendrán a μ , el parámetro.

El siguiente gráfico ilustra lo expresado anteriormente:



Con cada una de las medias, representados mediante su distribución muestral, se obtienen distintos intervalos de confianza, la mayor parte de ellos incluyen al verdadero valor del parámetro y una pequeña porción no lo contiene. En el caso del 95% de confianza, se espera que sólo el 5% de todos los intervalos construidos no incluyan al parámetro estimado.

En términos de clase lo recomendable es realizar una simulación, preferiblemente con apoyo del computador, para lograr una demostración empírica, potenciando la comprensión de los estudiantes de esta difícil abstracción y tratando de evitar que la explicación no se transforme en un simple acto de fe.

La demostración empírica de que la mayoría de los intervalos de confianza incluyen al parámetro estimado, es sólo una primera etapa, ya que lo cierto es que en la práctica sólo se dispone de una única repetición del experimento, por tanto de un único intervalo de confianza. Es importante recordar siempre al estudiante que en el enfoque frecuencial (el que se estudia en los cursos básicos de estadística) μ es una constante y que los límites del intervalo son aleatorios, por lo tanto, cuando se construye un intervalo de confianza sólo se tienen dos posibilidades:

- El intervalo contiene a la verdadera media poblacional, o
- El intervalo no contiene a la verdadera media poblacional. La muestra seleccionada es una de las pocas que no producen intervalos que incluyan al parámetro.

En conclusión, no se puede saber si el intervalo construido contiene o no el parámetro de interés, sin embargo, se tiene una *confianza* de que el parámetro desconocido se encuentre en él. Esa confianza se expresa mediante el *nivel de confianza*, así, para el 95% de confianza, se sabe que el intervalo fue construido por un método que da resultados correctos en el 95% de las veces. Se confía en que el intervalo construido a partir de la muestra aleatoria sea uno de los del grupo que contiene al parámetro estimado, aunque no se tiene la seguridad de que sea así; se sabe que se tiene un riesgo del 5% de estar equivocado.

De forma general, un intervalo de confianza de $100(1 - \alpha)\%$ puede interpretarse como si al tomar todas las muestras posibles de tamaño n , y se calcularan las medias muestrales y se construyeran los intervalos pertinentes, el $100(1 - \alpha)\%$ de esos intervalos construidos incluirán al verdadero valor de la media poblacional. La interpretación complementaria se expresa diciendo que se espera que solamente el $100\alpha\%$ de los intervalos construidos no contengan a la media poblacional. Una versión simplificada de esta interpretación para un intervalo específico con límites a y b es: se tiene una confianza del $100(1 - \alpha)\%$ de que la media poblacional se encuentre entre a y b .

Interpretaciones como: *La probabilidad de que el verdadero valor de la media poblacional se encuentre entre a y b es del $100(1 - \alpha)\%$ o existe una probabilidad del $100(1 - \alpha)\%$ de que la media poblacional se encuentre entre a y b* , son incorrectas desde la perspectiva del enfoque frecuencial. En estas interpretaciones se da a entender el parámetro es aleatorio y que el intervalo es fijo, por lo tanto hay una cierta probabilidad de que la media poblacional se encuentre entre esos límites. Esta interpretación puede deberse a una comprensión incorrecta de la expresión **A**, pero también se corresponde con la interpretación bayesiana.

El enfoque bayesiano se fundamenta en el Teorema de Bayes sobre el cálculo de probabilidades condicionadas y el parámetro que se desea estimar θ se considera una variable aleatoria. Otros puntos importantes de este enfoque son: todo lo que es desconocido es incierto y toda incertidumbre tiene que ser cuantificada en términos probabilísticos; existe una información inicial o a priori (anterior a la experiencia) que puede traducirse formalmente en una distribución de probabilidad para θ . En consecuencia, la distribución de probabilidad $p(\theta)$ es conocida y proporciona el grado de credibilidad que una persona atribuye a los posibles valores θ . Los datos recolectados servirán para actualizar la información ya existente sobre el parámetro. Luego se combina la información a priori con la información muestral y utilizando el Teorema de Bayes se atribuyen nuevos grados de credibilidad a los valores del parámetro, que dan lugar a la distribución a posteriori $p(\theta/x)$. A partir de esa distribución se puede calcular la probabilidad de que el parámetro esté en un intervalo, en consecuencia, la conclusión acerca del parámetro desconocido se expresa en términos de probabilidad. Por ejemplo: se tiene una probabilidad del 95% de que la verdadera media poblacional se encuentre entre a y b (Scotto y Tobias-Garcés, 2003). Obsérvese que la interpretación anterior es diferente a expresar que, en un muestreo repetido, el intervalo (a, b) contendrá a la media poblacional con una probabilidad del 95%. Para los bayesianos $100(1 - \alpha)\%$ es el nivel de credibilidad y los intervalos no son de confianza sino de credibilidad.

La investigación de los intervalos de confianza en lo que respecta a su enseñanza, aprendizaje y evaluación, todavía es incipiente y gran parte de los trabajos encontrados se han realizado con investigadores y autores de artículos publicados en revistas científicas. Por ejemplo, Belia, Fidler y Cumming (2005) trabajaron con investigadores que eran autores de artículos en revistas de psicología, neurociencia del comportamiento y medicina, encontrando que interpretaban el nivel de confianza como la probabilidad de que el intervalo de confianza contenga los valores de las medias muestrales. La mayoría de los investigadores con quienes trabajaron Cumming, Williams y Fidler (2004) no entienden lo que significa un intervalo de confianza. Los investigadores con que trabajaron Cumming et al (2004) adjudicaban una alta probabilidad a que al obtener una nueva muestra su media cayera de nuevo en el intervalo de confianza original. Consideran Cumming y sus colaboradores que este resultado es consistente con la denominada *Ley de los pequeños números* de las investigaciones de Tversky y Kahneman (1982).

Behar (2001) trabajó presentando las mismas preguntas a estudiantes y profesores de estadística y señala que la información que genera la comparación de las dos muestras estudiadas hace dudar sobre el supuesto de que los profesores tienen claros los conceptos que enseñan a sus estudiantes.

En cuanto a los estudiantes, al parecer, tienen problemas para establecer relación entre el tamaño del intervalo con el nivel de confianza y el tamaño muestral. Behar (2001) reporta que los estudiantes no establecen claramente la relación del nivel de confianza con el ancho del intervalo. Los resultados de Olivo y Batanero (2007) sugieren que los estudiantes presentan una amplia variedad de posibles dificultades conceptuales, procedimentales e interpretativas en relación al intervalo de confianza. Estos autores señalan que los estudiantes al parecer no logran comprender la relación entre precisión y tamaño de muestra o entre esta, la variación en los datos y el coeficiente de confianza. Behar y Yañez (2009) encontraron que tanto los estudiantes como los profesores y expertos poseen concepciones equivocadas respecto a los intervalos de confianza. Hay coincidencias en ambos grupos en cuanto a considerar que los intervalos contienen medias muestrales o valores individuales en lugar de posibles valores del parámetro. Asimismo la interpretación del nivel de confianza como me-

dida de certeza, sin ningún referente frecuencial, es habitual en los grupos estudiados por Behar y Yañez (2009).

Método

En la revisión bibliográfica realizada no se encontraron estudios sobre la interpretación de intervalos de confianza efectuados en Venezuela, por lo que se consideró pertinente realizar un estudio exploratorio. Hernández, Fernández y Baptista (2004) indican que las investigaciones de este tipo son adecuadas cuando el tema o problema a tratar se ha estudiado poco o nunca ha sido tratado en una comunidad pero se estima la presencia de un asunto social significativo. En este caso la investigación sobre la interpretación de intervalos de confianza por parte de estudiantes venezolanos permitirá obtener información para elevar el grado de familiaridad con el tema que se utilizará como base para el desarrollo de otras investigaciones.

El Instrumento

Para indagar sobre la comprensión que tienen los docentes en formación sobre los intervalos de confianza, se utilizó un cuestionario escrito de 14 ítems cerrados, del tipo Verdadero – Falso. Los ítems fueron diseñados por Behar (2001) con el propósito de identificar concepciones alternativas sobre intervalos de confianza, reportadas previamente en otras investigaciones. Se utilizaron 7 ítems con respuestas falsas y 7 con respuestas verdaderas respecto a intervalos de confianza.

En el cuestionario se identifican tres partes, una referida a una situación sobre pesos de escolares, para la interpretación de información de intervalos de confianza para la media en una población. La segunda parte hace referencia a intervalos de confianza para diferencias de medias en el contexto de la producción de leche de vaca. En las dos primeras partes el interés se centra en la naturaleza conceptual de los intervalos de confianza y su interpretación. En la última parte se presentan algunas afirmaciones sobre las relaciones que existen entre los diferentes elementos que conforman a los intervalos de confianza, como la desviación estándar de la población, el tamaño muestral y el nivel de confianza.

Los sujetos

Los participantes se seleccionaron entre un grupo de docentes en formación que cursaban la asignatura Estadística Aplicada a la Educación, que como su nombre sugiere, tiene como orientación la utilización de la estadística en procesos de investigación en educación. Lo usual es que en estos cursos se trabaje presentando los contenidos en contextos del área de la educación (notas, puntuaciones de un test, pesos o alturas de estudiantes, etc.). Para el momento de aplicación del cuestionario ya se había culminado el estudio de las distribuciones muestrales y de los intervalos de confianza para la media y la proporción. En esta asignatura el trabajo comienza con el estudio de la distribución muestral de medias, a partir de "poblaciones" de 4 o 5 elementos, con el propósito de evidenciar el cumplimiento del TCL y algunos ejemplos de aplicación. Luego se trabaja la estimación puntual y por intervalos para la media poblacional, sobre la base del conocimiento de la distribución muestral de medias y del TCL. Estos temas se trabajaron de forma tradicional, sin la utilización de recursos tecnológicos para hacer simulaciones. En uno de los cursos se realizaron construcciones de intervalos de confianza para la media a partir de muestras de distinto tamaño de una población de media conocida, de tal manera que los sujetos verificaran que el intervalo puede incluir o no al parámetro. Además, se les pidió calcular la proporción de intervalos que incluían al parámetro, considerando un mismo tamaño de muestra.

La aplicación del cuestionario se realizó sin previo aviso y como parte de una actividad regular del curso y fue respondido por todos los asistentes a la clase en las cinco secciones, tres diurnas y dos nocturnas, que se dictaban en ese momento. Por lo tanto, los participantes fueron seleccionados mediante un muestreo no probabilístico, del tipo estratégico o intencional, que de acuerdo con Azorín y Sánchez Crespo (1986) es el indicado para estudios exploratorios, ya que permite reducir costos y lograr mayor control del proceso. La muestra quedó conformada por 102 docentes en formación, con características semejantes en cuanto a la edad pero con un fuerte sesgo hacia el sexo femenino (70% de mujeres).

Resultados y análisis

Los resultados se presentan en tres partes, que coinciden con cada una de las partes del cuestionario descritas en el apartado anterior. Primero se expone la pregunta, tal como estaba en el cuestionario, acompañada de los porcentajes de estudiantes que señalaron cada alternativa, incluyendo los que optaron por no responder (respuestas en blanco); luego se realiza el análisis de las respuestas. Se destacan en negrilla los porcentajes de respuesta correctas.

Parte I

Para un conjunto de pesos (en kilogramos) de estudiantes de 3er. grado se construyó un intervalo del 95% de confianza para estimar la media, con base en una muestra aleatoria y resultó ser (42; 48). Para cada una de las siguientes afirmaciones, indique si es verdadera (V) o falsa (F), según su criterio.

Cuadro 1. Definición e interpretación de los Intervalos de Confianza para estimar la media poblacional

		V	F	B
1	95% de los pesos están entre 42 y 48 kilogramos. (F)	43,1	54,9	2,0
2	La mayoría de los pesos están entre 42 y 48 kilogramos. (F)	73,5	24,5	2,0
3	La probabilidad de que el intervalo incluya la media muestral es 95%. (F)	66,7	27,5	5,9
4	Si se generan 200 intervalos de confianza del mismo proceso aproximadamente 10 de aquellos intervalos no contendrán la media de la población. (V)	18,6	62,7	18,6
5	La probabilidad de que el intervalo (42,48) incluya la media poblacional es 95%. (F)	74,5	21,6	3,9
6	Un intervalo de confianza para una media siempre contiene la media de la muestra. (V)	62,7	30,4	6,9

En principio se destacan las afirmaciones 1 y 6, donde la mayoría de los docentes en formación encuestados responde de forma correcta. La primera de ellas asume que el intervalo de confianza contiene valores poblacionales y que el nivel de confianza denota el porcentaje de esos valores contenidos en el intervalo, mientras que la segunda asegura la presencia de la media muestral en el intervalo; pareciera entonces que la mayoría de los encuestados están claros en cuantos a estos aspectos. Sin embargo, el análisis de las otras respuestas evidencia otra cosa. El 73,5% de ellos piensa que un intervalo de confianza contiene la mayoría de los valores de la variable en estudio (afirmación 2). Aunque las proposiciones 1 y 2 son equivalentes, aunque expresada de forma distinta, los estudiantes participantes identifican una como falsa y otra como verdadera. Da la impresión que ellos consideran que los intervalos de confianza contienen valores poblacionales y no valores de la distribución muestral media muestrales al ser considerada como variable aleatoria.

Con las afirmaciones 3 y 6 también describen contradicciones de los encuestados. En la 6 identifican mayoritariamente que un intervalo de confianza siempre contiene a la media de la muestra, pero previamente en la 3 la mayoría había indicado que esa media se encontrará en el intervalo con una probabilidad del 95%. Mientras que en la 6 se da como un evento seguro, en la 3 se le ve como un evento probable, este tipo de respuesta puede deberse a una interpretación errónea de la expresión **A**, considerándola como centrado en μ y no en \bar{x} .

La afirmación 5, como se señaló antes, expresa una interpretación bayesiana del intervalo de confianza y aun cuando ellos sólo han trabajado la concepción frecuencial la consideran verdadera. Obviamente se trata de una confusión, ya que en esta concepción es erróneo pensar que el parámetro es aleatorio, mientras que el intervalo es fijo, como aparece en la afirmación 5.

La proposición 4 expresa una interpretación correcta del nivel de confianza en términos frecuenciales, expresada con relación al número de intervalos que se espera no contengan a la media poblacional. Sólo el 18,6% de los encuestados la consideran plausible, lo cual sin duda, ratifica las dificultades que ellos tienen en cuanto a la interpretación de los intervalos

de confianza, tal como lo señalan Olivo y Batanero (2007) y Behar y Yañez (2009). Llama la atención el porcentaje de respuestas en blanco en la afirmación 4. Mientras que en cada una de las otras proposiciones más del 90% de los encuestados piensan que tienen la respuesta correcta, en este caso, cerca del 19% de ellos prefiere no ofrecer respuesta alguna. Esto podría deberse a que los docentes en formación encuestados están poco familiarizados con este tipo de interpretación.

De acuerdo a los resultados de esta primera parte, los docentes en formación tienen dificultades en la definición e interpretación de los intervalos de confianza para estimar la media poblacional en un contexto que se supone cercano al encuestado.

Parte II

¿Tocar música a las vacas aumenta su producción de leche? Un experimento fue conducido con el propósito de responder esta pregunta y un conjunto de vacas lecheras fue dividido en forma aleatoria en dos grupos. Se tocó música a un grupo; al otro grupo (de control) no se le tocó música. El aumento medio en la producción de leche (Vacas con Música – Vacas sin música) fue 2,5 litros por vaca durante el periodo de tiempo de estudio. Un intervalo del 95% confianza para la diferencia de medias en la producción resultó ser (1,5; 3,5) litros por vaca. Responda verdadero (V) o falso (F), según su criterio, para las afirmaciones que se mencionan a continuación.

Cuadro 2. Interpretación de los Intervalos de Confianza para diferencias de medias

		V	F	B
7	El 95% de las vacas aumentaron su producción entre 1,5 y 3,5 litros por vaca. (F)	55,9%	43,1	1,0
8	Estamos 95% seguros que el aumento medio de la producción en la muestra es 2,5 litros por vaca. (F)	45,1%	52,0	2,9
9	Estamos 95% seguros que el aumento medio en la producción está entre 1,5 y 3,5 litros por vaca (V)	77,5%	19,6	2,9

La afirmación 7 es similar a la 1 y en ambos casos un porcentaje importante de los encuestados proporciona la respuesta correcta, pero a diferencia de 1, en este caso los que responde correctamente no son la mayoría. La afirmación 8 coloca el énfasis en la estimación puntual en una interpretación incorrecta de la estimación por intervalos. La mayoría de los encuestados la considera una interpretación inadecuada, aunque el porcentaje de sujetos que la consideran plausible no es despreciable (45%). Se puede afirmar entonces que la mayoría de los encuestados no entienden el nivel de confianza al estimador puntual de la media.

En la proposición 9 se presenta una interpretación correcta de los intervalos de confianza para diferencia de medias que es asumida como tal por la mayoría de los sujetos: 77,5%. Obsérvese que los porcentajes de aceptación de las proposiciones 5 y 9 son muy semejantes (74,5% y 77,5%), además de ser los más altos de todo el cuestionario. Ambas afirmaciones presentan de forma explícita el intervalo calculado, así como el nivel de confianza y hacen referencia directa al parámetro que se desea estimar. En ambos casos los estudiantes consideran verdaderas la interpretación, sólo que en 5 la interpretación es errónea mientras que en 9 es correcta. Los estudiantes parecen dar el mismo tratamiento a una interpretación subjetiva (5) como a una frecuencial (9), lo cual también se refleja en los altos porcentajes de encuestados que consideran verdaderas las proposiciones de esta parte del cuestionario. También podría suceder que los estudiantes identifican "probabilidad" con "estamos seguros".

Otro punto importante de destacar en esta parte del cuestionario, es el porcentaje de respuesta en blanco que no alcanza el 3% en ninguna de las afirmaciones, lo cual sugiere que la mayoría de los encuestados consideró que conocía la respuesta adecuada a cada proposición. Ese bajo porcentaje de respuestas en blanco podría deberse a que todas las afirmaciones de esta parte del cuestionario se formulan en la forma más común de interpretar los intervalos de confianza y no en la forma complementaria tal como aparece en la afirmación 4.

Los docentes en formación muestran un mejor desempeño en esta parte del cuestionario, que involucra intervalos de confianza para diferencias de medias. No obstante, los resultados de esta parte sugieren que el grupo encuestado asume el coeficiente de confianza como un valor de probabilidad en términos subjetivos, como un valor de certeza, sin ninguna interpretación frecuencial.

Parte III

Las afirmaciones de esta parte buscan indagar sobre la comprensión que tienen los encuestados respecto a las relaciones que existen entre los intervalos de confianza y diferentes elementos que los conforman, por lo tanto, también se indaga sobre la forma de construirlos: A continuación los resultados.

Cuadro 3. Relación entre el ancho de los Intervalos de Confianza y algunos conceptos que los definen

		V	F	B
10	Si se conserva el tamaño de la muestra fijo, el intervalo de confianza se pone más ancho cuando se aumenta el nivel de confianza. (V)	46,1	48,0	5,9
11	Si se conserva el nivel de confianza fijo, el intervalo de confianza se hace más estrecho cuando se aumenta el tamaño de la muestra. (V)	48,5	42,2	11,8
12	Si la desviación estándar de la población aumenta, el intervalo de confianza disminuye en la anchura (F)	42,2	41,2	16,7
13	El ancho del intervalo puede ser reducido al reducir el nivel de confianza (V)	56,9	33,3	9,8
14	El ancho del intervalo puede ser reducido al reducir la desviación estándar de la población (V)	37,3	47,1	15,7

En esta parte del cuestionario se destaca la proposición 13 por ser la única donde más del 50% las respuestas son acertadas, en el resto de las afirmaciones el total de buenas respuestas es inferior a la mitad de los encuestados. Los resultados sugieren que la mayoría de los encuestados reconoce que existe una relación directa entre reducir la amplitud del intervalo y reducir el nivel de confianza. No obstante, cuando se formula una proposición similar (10) pero en sentido contrario: al aumentar el nivel de confianza se amplía el intervalo, sólo el 46,1% considera que es cierta esta afirmación, un 10,8% menos de los que indicaron la proposición 13 como verdadera. Surge la duda si realmente la mayoría de los encuestados tienen clara la relación que existe entre la amplitud del intervalo y el nivel de confianza utilizado para su definición. La diferencia principal entre las afirmaciones 10 y 13 es la mención que se hace en 10 con relación a mantener el tamaño de muestra fijo, por lo que podría pensarse que este elemento confunde a los encuestados o que no reconocen el papel que juega el tamaño de la muestra con relación a la amplitud del intervalo y el nivel de confianza.

El tamaño de la muestra aparece también en la afirmación 11, en esta ocasión relacionándolo directamente con la amplitud del intervalo. Un 46% de los sujetos considera que la proposición es verdadera, sólo 4% más de los que la consideran falsa. Asimismo, se encuentra que un 11,8% de los sujetos tiene dudas respecto a la proposición por lo que prefiere dejarla en blanco; por lo tanto se puede inferir que la mayoría de los sujetos no tienen clara la relación existente entre la amplitud del intervalo y el tamaño de muestra utilizado para su construcción.

Las proposiciones 12 y 14 vinculan la amplitud del intervalo de confianza a la desviación estándar de la población, en el primer caso se logra un 41,2% de respuestas correctas, mientras que en el segundo sólo el 37,3%. En la afirmación 12, los porcentajes de sujetos que la consideran falsa (respuesta correcta) y los que creen que es verdadera es prácticamente el mismo, la diferencia es de 1% a favor de la respuesta incorrecta. Si se considera además que el 16,7% de los estudiantes optaron por no ofrecer respuesta, entonces se debe pensar que mayoría de los encuestados o bien

piensan erróneamente o tienen dudas sobre la relación que existe entre la amplitud del intervalo de confianza y la desviación estándar de la población. Esta apreciación se ratifica en la proposición 14, donde el 47% de los sujetos cree que la amplitud del intervalo no disminuye al reducir la desviación estándar de la población, y el 15,7% de ellos prefirió no tomar posición ante esta afirmación.

Pareciera entonces que los docentes en formación encuestados no tienen claro la relación que existe entre la amplitud de los intervalos de confianza y elementos como el tamaño de la muestra, la desviación estándar de la población y el nivel de confianza. La mayoría de ellos sólo reconoce el efecto que causa en la amplitud del intervalo cuando se reduce el nivel de confianza, no obstante, como grupo tienen dudas sobre qué pasa al aumentar ese nivel de confianza. Esas mismas dudas parecen presentarse cuando se indaga sobre la relación entre la amplitud del intervalo y el tamaño de la muestra o la desviación estándar de la población. Las inconsistencias que muestran los resultados podrían tener su origen en una comprensión parcial o inadecuada del concepto de distribución de medias muestrales, el TCL y su relación con la construcción de los intervalos de confianza.

A manera de conclusión

De acuerdo con el análisis de los resultados, los docentes en formación, que participaron en este estudio, interpretan de forma errónea los intervalos de confianza. Los resultados sugieren que tienen dificultades en la definición e interpretación de los intervalos de confianza para la estimación de la media poblacional en los dos contextos presentados. Los sujetos encuestados, al igual que los de Cumming, Williams, y Fidler, (2004), parecen ignorar que los intervalos de confianza se construyen a partir de muestras y que por lo tanto sus límites son aleatorios, en consecuencia no se puede aseverar que un intervalo en particular contiene al parámetro estimado. Los encuestados proporcionan evidencias que llevan a pensar que interpretan los intervalos de confianza en correspondencia con la concepción bayesiana y no en la tónica frecuentista como se supone estudiaron.

La mayoría de los encuestados sólo es capaz de reconocer el efecto de la reducción del nivel de confianza en la amplitud del intervalo, pero el grupo se presenta dividido cuando se indaga sobre el efecto de aumentar el nivel de confianza. Al indagar sobre el efecto de cambios del tamaño de la muestra o la desviación estándar de la población en la amplitud de los intervalos de confianza, el grupo también se muestra dividido. Parece que los docentes en formación encuestados tienen dificultades en la comprensión de la relación de la amplitud de los intervalos de confianza con el tamaño de la muestra, el nivel de confianza o la desviación estándar de la población. Estos resultados son semejantes a los de Behar y Yañez (2009).

En general, el grupo no fue consistente en sus respuestas, cuando la mayoría identificó una respuesta correcta en una proposición, luego una parte importante de ellos proporcionaba la respuesta incorrecta en otra proposición que evaluaba el mismo tópico pero con una presentación diferente. Los docentes en formación interpretan de forma errónea la información proporcionada por los intervalos de confianza y presentan serias dudas en cuanto a su definición, así como la relación de ellos con algunos de los elementos que se utilizan en su construcción.

Varios de los errores aquí identificados pueden tener su origen en una comprensión parcial o inadecuada del concepto de distribución de medias muestrales, el TCL y su relación con la construcción de los intervalos de confianza. Parecería entonces necesario desarrollar investigaciones que permitiere despejar o ratificar esta apreciación. Habría que investigar si un tratamiento más detallado de las distribuciones muestrales y del TCL ayuda a una mejor comprensión de la interpretación frecuencial de los intervalos de confianza. En este sentido existen recomendaciones de trabajar más con simulaciones, con programas como *Fathom Dynamic Data*, ya que permiten que los estudiantes puedan involucrarse directamente con la construcción de las distribuciones muestrales y en la comprobación del TCL, en lugar de presentar solamente el resultado final, como suele ocurrir en los cursos tradicionales de estadística.

Asimismo sería conveniente indagar sobre la forma como interpretan los intervalos de confianza los profesores de los cursos de donde se seleccionaron los participantes de este estudio, además, de la secuencia didáctica seguida para este tema. Algunos de los errores, como la interpretación bayesiana de los intervalos de confianza, pueden ser producto de la enseñanza o consecuencia de conceptos poco claros de parte de los profesores de estadística (Behar, 2001).

REFERENCIAS

Azorin, F. y Sánchez Crespo, J. L. (1986). *Métodos y aplicaciones del muestreo*. Madrid: Alianza.

Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística*. Tesis doctoral no publicada. Universitat Politècnica de Catalunya. Barcelona, España.

Behar, R. y Yañez, G. (2009). *Concepciones que los Expertos y los Estudiantes tienen respecto a los Intervalos de Confianza*. Ponencia presentada en: 2º Encuentro Iberoamericano de Biometría. V Reunión de la Región Centroamericana y del Caribe de la Sociedad Internacional de Biometría. 27 al 31 de julio de 2009/World Trade Center, Boca de Río, Veracruz, México.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). *Researchers misunderstand confidence intervals and standard error bars*. *Psychological Methods*, 10, 389 – 396.

Cohen, J. (1994). *The earth is round ($p < .05$)*. *American Psychologist* 49 (12), 997 – 1003.

Cumming G, Williams J, Fidler F. (2004). *Replication and researchers' understanding of confidence intervals and standard error bars*. *Understanding Statistics*, 3, 299 – 311.

Davies, H y Crombie, I.K. (2009): *What are confidence intervals and p-values?* Disponible en: <http://www.medicine.ox.ac.uk/bandolier/painres/download>

/whatis/What_are_Conf_Inter.pdf

Glass, G. y Stanley, J. S. (1994). *Métodos Estadísticos Aplicados a las Ciencias Sociales*, Ed. Prentice-Hall Hispanoamericana, México.

Hernández R., Fernández C. y Baptista L. (2004). *Metodología de la Investigación*. México: Mc Graw Hill.

Johnson, R. y Kuby, P. (2004). *Estadística Elemental*. México: Thompson.

Killeen P. R. (2006) *Replicability, Confidence, and Priors*. Psychological Science, 16(12), 1009 – 1012.

Olivo, E. y Batanero, C. (2007). *Un estudio exploratorio de dificultades de comprensión del intervalo de confianza*. Unión, 12, 37 – 51.

Scotto, M.G y Tobias-Garces, A. (2003) *Interpretando correctamente en salud pública estimaciones puntuales, intervalos de confianza y contrastes de hipótesis*. Salud pública de México, 45 (6), pp. 506 – 511.

Thompson, B. (2002). *What Future Quantitative Social Science Research Could Look Like: Confidence Intervals for Effect Sizes*. Educational Researcher, 31 (3), 25 – 32.

Tversky, A, y Kahneman, D. (1982). *Belief in the law of small numbers*. En: Kahneman, D., Slovic, P. y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases* (p. 23-31). New York: Cambridge University Press.

Wilkinson, L. (1999). *Statistical methods in psychology journals: Guidelines and explanations*. American Psychologist, 54, 594 – 604.

Autores:

Aurora LaCueva

María de los Ángeles Martín

Oswaldo J. Martínez P

María S. Harrington M

Ana C. Bolívar O.

Mildred C. Meza Ch

María Amelia Migueles

Juan C. Álvarez

María E. D' Aubeterre

Tulio Ramírez

Mariángeles Páyer Sánchez

María Gorety Rodríguez

Carlos E. Rojas C

Evelina Tineo D

Amalio Sarco Lira

Ramón Uzcátegui

Luis Bravo J

Evelyn C. Tineo

Zuly Millán

Carmen Ponte

Audy Salcedo

Jesús González

Gabriel Yañez

ISBN: 978-980-00-2694-6



9 789800 026946

Universidad Central de Venezuela.

Facultad de Humanidades y Educación. Escuela de Educación.

Centro de Investigaciones Educativas (CIES). Caracas - Venezuela