

Una demostración en el aula para ayudar a los estudiantes a entender los intervalos de confianza.

Christine M. Anderson-Cook
Tecnológico de Virginia

Journal of Statistics Education v.7, n.3 (1999)

Derechos de autor (c) 1999 Christine M. Anderson-Cook, todos los derechos reservados. Este texto se puede compartir libremente entre personas, pero no puede ser reproducido en cualquier medio sin consentimiento escrito expreso de la notificación del autor y del aval del editor.

Palabras claves: Aprendizaje activo; Cursos introductorios de la estadística; Intervalos puramente subjetivos.

Resumen

Este artículo discute una técnica de aprendizaje activo que se puede incorporar fácilmente a una diversidad de clases introductorias de estadística para explicar intervalos puramente subjetivos e intervalos de confianza estadísticos. Los conceptos de intervalo de confianza, nivel de confianza, y parámetros poblacionales fijos aunque desconocidos, son con frecuencia entendidos incorrectamente por un porcentaje significativo de estudiantes. La actividad en clase aquí presentada permite presentar estos conceptos, haciendo énfasis en la naturaleza objetiva de los intervalos estadísticos de confianza. También se hace hincapié en que la precisión del intervalo depende de la calidad de los datos usados en su construcción. El ejercicio propuesto toma menos de 50 minutos de tiempo de la clase y ayuda a fortalecer estos conceptos estadísticos esenciales de una manera gráfica y fácil de recordar. La reacción de los estudiantes al ejercicio ha sido positiva según lo medido en la práctica con base en la comprensión mejorada de los conceptos y el interés creciente en la actividad.

1. Introducción

Este ejercicio se utiliza actualmente en cursos introductorios de estadística aplicada en una diversidad de disciplinas, incluyendo la ingeniería, las matemáticas, y las ciencias biológicas, físicas y sociales. El propósito del ejercicio es presentar visualmente el concepto de nivel de confianza y la naturaleza fija, aunque desconocida, del parámetro poblacional que se está estimando. La actividad pone énfasis en la naturaleza objetiva del intervalo de confianza y en cómo la precisión del intervalo depende de la calidad de la información obtenida con estrategias de muestreo apropiadas.

2. Una característica clave del ejercicio es que los estudiantes participan activamente en la demostración de los conceptos, lo cual es esencial para el incremento del aprendizaje del estudiante ([Angelo 1993](#)). Esta interacción con los estudiantes y su involucramiento en todas las fases de la actividad ayudan a consolidar conceptos estadísticos difíciles.

3. La actividad consiste de tres fases. En la fase 1, los estudiantes proporcionan sin mayor preparación “un intervalo de confianza del 95%” puramente subjetivo para un hecho sobre el que tienen cierta intuición pero, probablemente, ningún conocimiento preciso. Se presenta un diagrama de sus intervalos, y se realiza una discusión sobre lo que mide ese “95%”. En la fase 2, los estudiantes emplean, para un hecho diferente, el mismo método puramente subjetivo en la construcción del intervalo, pero ahora deben tener una mejor comprensión de las condiciones que se espera que satisfagan los intervalos. En ambas fases, los estudiantes tienden a hacer los intervalos más estrechos y, por lo tanto, éstos no incluyen el parámetro verdadero con la frecuencia apropiada. En la fase 3 hacen la transición de un intervalo puramente subjetivo a un intervalo de confianza estadístico. En esta fase, los estudiantes examinan las semejanzas y las diferencias entre los dos tipos de intervalos. La fase concluye con una demostración de la construcción de intervalos de confianza para la media de una población con distribución normal usando muestras aleatorias simples. Esta fase puede

también incluir una discusión de los supuestos distribucionales y de los requerimientos de tamaño de muestra para el uso de los intervalos de confianza estadísticos.

4. El ejercicio toma menos del tiempo de una clase (uso aproximadamente 40 minutos para llevar a cabo todas las demostraciones y la discusión), y requiere de la participación activa de los estudiantes. El tamaño ideal de la clase para la actividad es de alrededor de 40 personas, pero funciona para clases de 20 hasta de 100 estudiantes. Si la clase es demasiado pequeña, puede ser difícil ilustrar con eficacia el nivel de confianza del 95%. Si la clase es mucho mayor de 40 personas, sólo debe utilizarse un subconjunto de estudiantes en cada una de las fases 1 y 2.

5. Los materiales requeridos incluyen una hoja de datos estadísticos poblacionales ([tabla 1](#)), dos o tres tarjetas de archivo o pedazos pequeños de papel para cada estudiante de la clase, y un proyector de acetatos con varias transparencias (una transparencia hecha de una fotocopia de papel para gráficas resulta adecuado). Alternativamente, las transparencias se pueden substituir por una pantalla de proyección y una computadora portátil, si se incorporan las funciones de S-Plus en el [apéndice 1](#) en la demostración. Este segundo acercamiento permite una presentación levemente más rápida de los datos, pero ambos métodos funcionan bien.

Tabla 1. Hechos a estimar con intervalos de confianza (basados en datos de 1996)

Categoría	Hecho	Unidades	Valor verdadero
Estados Unidos	Área geográfica	Millas cuadradas	3,787,428
	Número de estudiantes de (educación básica, media y universidad)	Personas (en millones)	59.9 millones
	% de población en ciudades	Porcentaje	77.5
	Crecimiento poblacional de 1986 a 1996	Personas	22,164,068
	Densidad demográfica promedio	Personas por milla cuadrada	70.3
	Ingreso per cápita PIB	US\$ por persona	22,568
	Mano de obra total	Personas (en millones)	118.9 millones
Mundo	Área total de la tierra	Millas cuadradas	57.49 millones
	Altura del montaje Everest	Pies	29,030
	Ingreso per cápita PIB	US\$ por persona	5,900
	Densidad demográfica promedio	Personas por milla cuadrada	100.4

2. Detalles de la actividad

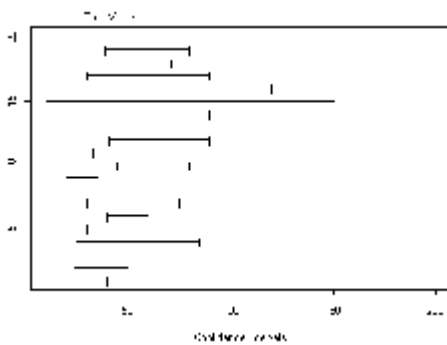
6. Esta actividad se desarrolla comúnmente al principio de la unidad dedicada a estimación, pero después de que los estudiantes hayan tenido una introducción a la probabilidad y a la distribución normal. Es útil haber tenido una discusión sobre el hecho de que los resultados de las encuestas generalmente se divulgan con un margen de error y un nivel de confianza (por ejemplo, el $12\% \pm 2\%$, correcto 19 veces de 20).

2.1 Fase 1

7. En la Fase 1, se les pide a los estudiantes que proporcionen los intervalos puramente subjetivos de confianza del 95% para un hecho dado (seleccionado de la columna de hechos presentados en la [Tabla 1](#)). Además de los hechos presentados en la Tabla 1, algunos hechos locales o estatales pueden resultar también de interés. Éstos están disponibles en informes anuales o en Internet. En general, se dan pocas explicaciones para aclarar qué es lo que significa un intervalo de la confianza del 95%; en su lugar, se deja a los estudiantes hacer sus propios juicios. Es provechoso dar a los estudiantes alguna escala para sus estimaciones (por ejemplo, “en millares, dar un intervalo de la confianza del 95% para...”).

8. Los intervalos se registran en las tarjetas de archivo, se recopilan, y son resumidos por un estudiante (generalmente un voluntario de la clase). Mientras que este estudiante está llenando los datos en la transparencia cuadriculada, puede discutirse con los estudiantes las ideas en las que se basaron para determinar la anchura de sus intervalos (es decir, sus nociones de “95% de confianza”). Es también provechoso hacer una discusión sobre qué constituiría el intervalo ideal (a saber, uno que sea estrecho, pero también correcto). Si se utilizan transparencias, es útil determinar un límite superior e inferior que incluya los intervalos de todos los estudiantes en la clase antes de que se recojan las tarjetas. Una vez que todos los intervalos se registran en la transparencia, se agrega una línea vertical para demostrar el valor verdadero del hecho bajo consideración. Alternativamente esto se puede hacer usando la función “conf.int” dada en el [Apéndice 1](#). [La Figura 1](#) muestra un ejemplo de cómo se vería el resultado para 20 estudiantes después de finalizar la fase 1.

9. Es típico que mucho más del 5% previsto de los intervalos de los estudiantes no incluyan el valor verdadero en su rango. En el ejemplo dado, solamente 11 de 20 estudiantes dieron los intervalos que incluyen el valor verdadero del parámetro. Esto puede dar pie a la discusión sobre lo que esto significaría en la práctica, si citamos un I. C. de 95%, pero el nivel de confianza verdadero real era mucho más bajo.



[Figura 1 \(4.0K gif\)](#)

Cuadro 1. Intervalos del estudiante para la fase 1.

2.2 Fase 2

10. En la fase 2, repítase la actividad de la fase 1 con una ligera modificación. Pida que cada estudiante dé un intervalo de confianza del 95% y proporcione solamente una única mejor predicción para un nuevo hecho. Registre las estimaciones puntuales y de intervalo. La finalidad de esta fase es ayudar a los estudiantes a formalizar algunos de los conceptos en términos de “sentido común”, antes de cambiar al dominio de lo estadístico, donde muchos estudiantes parecen menos dispuestos a confiar en su intuición.

11. La discusión de los intervalos de la Fase 2 da oportunidad de destacar resultados específicos de los intervalos construidos y de las características generales de los intervalos de confianza. Después de calcular la proporción de intervalos que contienen el valor verdadero, los estudiantes pueden determinar la precisión del

grupo de intervalos de confianza. Es muy frecuente que la mayoría de los estudiantes todavía haga los intervalos demasiado estrechos para incluir el valor verdadero, de manera que se puede no estar cerca del nivel de confianza deseado del 95%. También puede motivarse a los estudiantes a que consideren cómo los resultados y la deseabilidad de los intervalos cambiarían si el nivel deseado de confianza se cambia al 90% o al 99%. Por ejemplo, el nivel de confianza del 90% permite un intervalo más estrecho, pues hay menos presión de acertar que en el caso de los intervalos de confianza del 95% o del 99%. El examinar las estimaciones puntuales puede dar pie a la discusión de las ventajas de tener un rango de éxito cuantificable para los intervalos, contra la simplicidad de un solo valor para la estimación puntual. Finalmente, pueden discutirse los méritos de tener centrada la estimación puntual en el intervalo.

2.3 Fase 3

12. La Fase 3 se ocupa de hacer la transición de intervalos puramente subjetivos a los intervalos de confianza estadísticos. Primero, se deben hacer algunas observaciones sobre las características comunes de todas las estimados (*estimates*) de intervalos. La motivación para crear un intervalo es proporcionar un rango razonable de valores que tenga una alta probabilidad de incluir el parámetro verdadero en cuestión. El intervalo ideal será estrecho y preciso (contiene el valor verdadero). Debe quedarles claro a los estudiantes que el incrementar la cantidad de conocimiento usada para construir el intervalo debe conducir a un intervalo más estrecho, más preciso. Una vez más es provechoso destacar la carencia de cuantificación acerca de si las estimaciones puntuales están “cerca” del valor verdadero, y la forma en que los intervalos de confianza superan esta debilidad. Debe hacerse hincapié en que el nivel de confianza está basado en el muestreo repetido (tal como la construcción de muchos intervalos por parte de los miembros de la clase), pero para un caso específico, tenemos generalmente solamente un único intervalo bajo consideración. El intervalo de confianza único es correcto o incorrecto (todo o nada), pero el nivel de confianza nos da una indicación de la proporción de los intervalos correctos que se pueden esperar con la repetición del procedimiento de estimación. Finalmente, una vez que se construye un intervalo, generalmente no descubrimos si es en realidad correcto. Es importante hacer énfasis en esta cuestión final, pues ayuda a los estudiantes a centrarse en la diferencia entre la muestra, de la que están disponibles todos los datos, y la población, de la que comúnmente se desconoce la respuesta correcta.

13. En seguida se dedica algún tiempo a formalizar la conexión entre los intervalos puramente subjetivos y los intervalos de confianza estadísticos, así como a destacar las diferencias claves entre los dos métodos. Una muestra de la población es similar al conocimiento parcial usado para estimar la cantidad en cuestión. Tanto en los casos puramente subjetivos como en los estadísticos, disponemos de cierta información que nos ayuda a dar un rango de valores razonables, pero la carencia de conocimiento completo conduce a una cierta incertidumbre en las estimaciones. Esperamos que nuestro conocimiento nos permita hacer una estimación razonable, pero no tenemos seguridad de su exactitud. La cantidad de conocimiento usado para estimar el intervalo puramente subjetivo es más o menos análoga a la cantidad de datos recogida para un intervalo de confianza estadístico. Entre más datos o información estén disponibles, el intervalo se hace más estrecho. Para un intervalo puramente subjetivo, la información es medida por el nivel del conocimiento sobre el tema. Para un intervalo de confianza estadístico, la anchura depende del número de observaciones de la población en cuestión de la muestra.

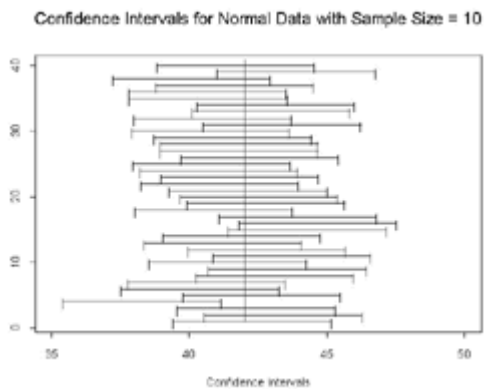
14. También puede compararse la calidad de la información utilizada para construir los intervalos y la precisión de estos. Si alguien sin el conocimiento exacto del tema sugiere un intervalo de confianza puramente subjetivo, no esperaríamos que su intervalo fuera útil para estimar la cantidad verdadera en cuestión. En forma similar, si una muestra no es representativa de la población subyacente que supuestamente describe, no deberíamos esperar que el intervalo proporcione una estimación exacta del valor de población.

15. La estimación de intervalos de varios estudiantes es más o menos análoga a la recolección de muestras múltiples. Cada estudiante o muestra produce un intervalo, pero es típico que en la práctica sólo se forme un

único intervalo, y que no sepamos si el intervalo es correcto. Una diferencia clave entre los métodos es que, dada la misma cantidad de conocimiento (por ejemplo, dos estudiantes que tomaron el mismo curso de geografía), es poco probable que los dos intervalos puramente subjetivos sean iguales. Por el contrario, dados los mismos datos, se obtiene un solo intervalo de confianza estadístico. Esta ausencia de subjetividad en el método estadístico es una ventaja significativa.

16. Otra diferencia entre los métodos es que, a menos que se pueda determinar el valor verdadero del parámetro en cuestión, no podemos cuantificar el nivel de confianza de un intervalo puramente subjetivo. Sin embargo, para un intervalo de confianza estadístico, existe la teoría subyacente que (bajo ciertas suposiciones distribucionales) nos da un nivel exacto de confianza para nuestros intervalos.

17. Hay dos opciones para concluir esta fase del ejercicio. Imprimir las figuras en el [Apéndice 2](#) en transparencias o utilizar las funciones de S-Plus del [Apéndice 1](#) para mostrar los resultados de los intervalos de confianza estadísticos. La función de S-Plus llamada “norm.conf” genera una gran cantidad de muestras de una distribución normal dada, y calcula el intervalo de confianza para la media de cada muestra. Un ejemplo de los resultados que arroja la función se da en la [Figura 2](#), para 40 muestras de tamaño 10 de una distribución normal con media de 42 y desviación estándar de 4 unidades. Alternativamente, uno puede utilizar el cuadro 2, junto con los cuadros [3](#) y [4](#) del [Apéndice 2](#), para demostrar una variedad de resultados. Estos diagramas muestran 40 intervalos generados usando diferentes tamaños de muestra y niveles de confianza.



[Figura 2 \(29.6K.jpg\)](#)

Figura 2. Intervalos del confianza del 95% generados por una computadora para estimar la media de una distribución normal, $n = 10$.

18. Estas figuras se pueden utilizar para discutir varias características claves de los intervalos de confianza estadísticos. Podemos evaluar el nivel de confianza verdadero de los intervalos de confianza observando la proporción de intervalos que contienen el parámetro verdadero de la población. En el ejemplo dado en la [Figura 2](#), 39 de los 40 intervalos de confianza contienen el valor verdadero del parámetro. Esto es consistente con datos de una distribución binomial con la probabilidad de éxito igual a 0.95. Las diferencias observadas en la anchura de los intervalos para un tamaño de muestra fijo son un reflejo de la calidad de los datos de las observaciones de cada muestra. El aumento del tamaño de la muestra da lugar a intervalos de confianza más estrechos, como puede observarse en las figuras [2](#) y [3](#). Disminuir el nivel de confianza reduce la anchura de los intervalos de confianza, como se muestra en las figuras [2](#) y [4](#). Finalmente, puede llevarse a cabo una discusión sobre cómo las violaciones de supuestos distribucionales pueden conducir a niveles de confianza incorrectos. Al final de este ejercicio, los estudiantes tienen la base conceptual y la motivación para aprender los detalles de la construcción de intervalos de confianza estadísticos para la media de una distribución normal.

3. Conclusiones

19. Este sencillo ejercicio se ha puesto en ejecución con éxito en varios cursos introductorios de estadística. Los estudiantes aprecian la demostración gráfica de los conceptos que han sido con frecuencia una fuente de confusión para ellos, y pueden discutir mejor el significado de los intervalos de confianza estadísticos, los niveles de confianza, y las ventajas del intervalo de estimación sobre la estimación puntual. La retroalimentación recibida de los estudiantes sobre esta actividad ha sido de aceptación. Al final del curso, muchos todavía recuerdan el ejercicio y comentan positivamente respecto a él. En los exámenes parciales y en el final, las preguntas sobre intervalos de confianza son contestadas correctamente en forma más consistente que antes, y los estudiantes pueden hablar con más confianza de estos conceptos. Se requieren relativamente pocos recursos especializados y solamente una cantidad de tiempo pequeña para introducir con eficacia este concepto importante y para construir un conocimiento estadístico creciente.

Reconocimientos

La investigación para este artículo fue financiada por el Centro para la Excelencia en la Enseñanza de Estudiantes de Licenciatura del Tecnológico de Virginia. También quisiera expresar mi agradecimiento a Tom Wonnacott, del Departamento de Ciencias Estadísticas y Actuariales, Universidad de Ontario Occidental, por la idea inicial de construir intervalos de confianza puramente subjetivos para demostrar este concepto. Además, quisiera agradecer a los revisores y al redactor asociado por sus comentarios, que ayudaron a mejorar substancialmente la legibilidad y el contenido de artículo.

Apéndice 1: Funciones S-Plus

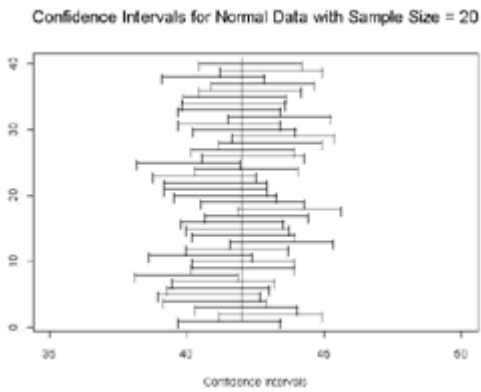
Función A.1 Para Trazar Intervalos Subjetivos Generados por los Estudiantes para un Hecho con Valor Verdadero “actual.val”

```
conf.int <- function(actual.val) { data <- scan() matrix.data <- matrix(data, byrow="T," ncol="2)" num.data <- length(matrix.data[, 1]) min.data <- min(matrix.data[, 1], actual.val) max.data <- max(matrix.data[, 2], actual.val) win.graph() plot(matrix.data[, 2], c(1:num.data), xlim="c(min.data," max.data), ylim="c(1," num.data), pch="| " , xlab="(" "Confidence Intervals)", ylab="(" " ") points(matrix.data[, 1], c(1:num.data), pch="|" ) for(i in 1:num.data) { lines(matrix.data[i, ], c(i, i)) } matrix.act <- matrix(actual.val, nrow="num.data" + 1, ncol="1)" lines(matrix.act, c(0:num.data + 1), lty="4") }
```

Función A.2 Para Generar Intervalos de Confianza para las Muestras de una Distribución Normal con tamaño de muestra, “num,” media “mu,” y desviación estándar “stdev”

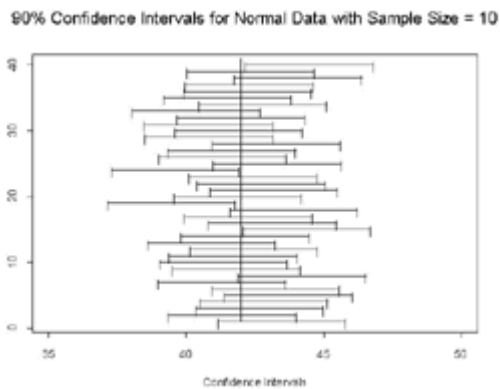
```
norm.conf <- function(num, mu, stdev) { num.sample <- 40 matrix.sample <- matrix(0, nrow="num.sample," ncol="2)" for(i in 1:num.sample) { sample <- rnorm(num, mu, stdev) mean.sample <- mean(sample) var.sample <- var(sample) lower.tquant <- mean.sample qt(0.975, num 1) * ( stdev/sqrt(num)) upper.tquant <- mean.sample + qt(0.975, num 1) * ( stdev/sqrt(num)) matrix.sample[i, 1] <- lower.tquant matrix.sample[i, 2] <- upper.tquant } # plot the graphs min.sample <- min(matrix.sample[, 1]) max.sample <- max(matrix.sample[, 2]) win.graph() plot(matrix.sample[, 2], c(1:num.sample), xlim="c(min.sample," max.sample), ylim="c(1," num.sample), pch="| " , xlab="("Confidence" Intervals)", ylab="(" " ") points(matrix.sample[, 1], c(1:num.sample), pch="|" ) for(i in 1:num.sample) { lines(matrix.sample[i, ], c(i, i)) } matrix.act <- matrix(mu, nrow="num.sample" + 1, ncol="1)" lines(matrix.act, c(0:num.sample + 1)) }
```

Apéndice 2: Diagramas adicionales



[Figura 3 \(jpg 27.2K jpg\)](#)

Figura 3. Intervalos de confianza del 95% generados en computadora para la media de una distribución normal, $n = 20$.



[Figura 4 \(28.2K jpg\)](#)

Figura 4. Intervalos de confianza del 90% generados por computadora para la media de una distribución normal, $n = 10$.

Referencia

Angelo, T. A. (1993), "A Teacher's Dozen: Fourteen General Research-Based Principles for Improving Higher Learning in Our Classrooms," *American Association of Higher Education Bulletin*, 45, 3-13.

**Christine M. Anderson-Cook
Departamento de Estadística
Tech de Virginia
Blacksburg, VA 24061-0439
*candcook@vt.edu***
