Construcción del concepto de intervalo de confianza mediante simulación en R

Construction of the confidence interval concept through simulation in R

Manuel Ricardo Contento Rubio

Universidad Nacional de Colombia

Facultad de Ciencias.

Bogotá, Colombia

2012

Construcción del concepto de intervalo de confianza mediante simulación en R

Construction of the confidence interval concept through simulation in R

Manuel Ricardo Contento Rubio

Tesis o trabajo de investigación presentado como requisito parcial para optar al título de:

Magister en Enseñanza de las Ciencias Exactas y Naturales

Director:
PhD Ramón Giraldo.

Universidad Nacional de Colombia
Facultad de Ciencias
Bogotá, Colombia

2012

A Martha, Nicolás y Gabriela.
Por entender la ausencia.

A mi padre por enseñar con su ejemplo.

Resumen

La estimación de parámetros a través de intervalos de confianza es una de las herramientas básicas del análisis estadístico. Esta temática se aborda en los curso de fundamentación estadística de los programas de pregrado e incluso en los programas de fundamentación media de algunos países. En este trabajo se presenta una propuesta pedagógica para su enseñanza basada en el uso de la simulación en R (http://www.r-project.org/). Por simplicidad se trabaja el caso de estimación por intervalo para la media (μ) asumiendo muestras de distribuciones binomial y normal.

Palabras clave: Intervalos de confianza, simulación, enseñanza

Abstract

The estimation of parameters across confidence intervals is one of the basic tools of the statistical analysis. This subject matter approaches in the course of statistical foundation of university and even in the programs of basic education of some countries. In this work one presents a pedagogic offer for his education based on the use of the simulation in R (http: // www.r-project.org/). For simplicity one works the case of estimation for interval for the average (μ) assuming samples of binomial and normal distributions.

Keywords: confidence Intervals, simulation, education

Contenido

Pág.

Lista de figuras

Pág.

Introducción

La estadística como disciplina es relativamente joven (Yañez, 2000) en comparación con la matemática, que en el siglo III a.c inició la formalización mediante un sistema axiomático que fue recopilado en el tratado *Los Elementos* (Euclides, trad. 1991), escrito por Euclides de Alejandría, considerado el matemático mas prominente de la antigüedad. Aunque los orígenes de la estadística son también antiguos y se remontan a actividades principalmente relacionadas con el conteo de bienes, productos y al manejo de las cifras del estado, de donde se acuño su nombre, es reciente su reconocimiento como disciplina relevante y de importancia práctica; gran razón tiene el eminente estadístico hindú Radhakrishna Rao al afirmar que "la estadística tiene gran antigüedad pero escasa historia" (Rao, 1997).

Dentro de la lista de autores que más han contribuido a la estadística se destacan algunos que han sido los precursores, cuyos aportes proporcionaron impulso a esta ciencia, entre ellos tenemos a Adolphe Quetelet (1796-1874), Francis Galton (1822-1911), Karl Pearson (1857-1936), Charles Spearman (1863-1945), William Gossett (1876-1937), Ronald Fisher (1890-1962), Jerzy Neyman (1894-1981), y Andréi Kolmogorov (1903-1987). Estos y muchos otros intelectuales además de su trabajo como investigadores, se ocuparon también de difundir su conocimiento en eventos científicos y en instituciones de educación; pero la problemática en torno a la enseñanza y el aprendizaje de la estadística tiene los primeros referentes significativos tan solo al final del siglo pasado: (Kempthorne, 1980; Garfield & Ahlgren, 1988).

Uno de los contenidos temáticos fundamentales en la enseñanza de la estadística es el análisis inferencial, que en términos muy generales consiste en estimar las características generales de las poblaciones mediante muestras. Su aplicación es compleja e involucra una gran cantidad de conceptos entre los que se destacan: la distribución de muestreo, las nociones de intervalos de confianza (IC) y las pruebas de hipótesis (PH), los cuales son fundamentales al momento de entender y aplicar con éxito los métodos estadísticos. Particularmente el aprendizaje del concepto de IC ha generado diversas tensiones entre estudiantes y profesores, debido a la complejidad e intrincadas relaciones entre los elementos constitutivos de esta noción. En el desarrollo histórico de la inferencia estadística y específicamente en la noción de IC se evidenciaron dificultades y controversias que se dieron lugar entre investigadores consumados que trabajaron en este campo, para finalmente proponer su actual estado.

Entender y reflexionar sobre el proceso de inferencia estadística favorece el aprendizaje de los futuros profesionales, particularmente de aquellos que serán usuarios frecuentes de los métodos estadísticos y por lo tanto requieren de conocimientos bien fundados que le permitan utilizar la estadística con criterio y responsabilidad. Aunque subsisten las dificultades, los docentes han hecho uso de diversas estrategias para abordar la enseñanza de la inferencia; particularmente la comprensión de la noción de intervalo de confianza ha sido objeto de estudio por parte de investigadores logrando aportes teóricos y filosóficos, como por docentes que han usado experiencias de aula como elemento constitutivo de sus indagaciones. Lo anterior invita a profundizar en el complejo estudio de los problemas de la enseñanza y aprendizaje de la estadística y a desarrollar estrategias para buscar soluciones en las mencionadas áreas.

Este trabajo tiene como aristas más visibles la revisión de la teoría relacionada con el concepto estadístico de intervalo de confianza, la exploración del estado del arte sobre las dificultades en su comprensión, el examen de documentación sobre el uso de herramientas computacionales para apoyar la enseñanza de este concepto; esto con el fin de estructurar una propuesta para enseñar IC usando simulación, recurso que ha ganado ímpetu, con el cual se espera apoyar el proceso de enseñanza y potenciar la comprensión en los estudiantes. El software usado para desarrollar las simulaciones es R (http://www.r-project.org/) el cual es libre y debido a que requiere elementos de programación, invita a los estudiantes a tomar parte activa en su proceso de aprendizaje.

En este documento se consigna en el primer capítulo una revisión sobre la evolución del concepto de IC, que se relaciona con los aspectos disciplinares y en particular con la inferencia estadística. Luego en el segundo capítulo se incluye la presentación de los problemas que se suscitan en la interpretación de los IC apoyándose en referencias e investigaciones al respecto así como en la experiencia docente; al final de este capítulo se hace una breve presentación sobre el uso de programas computacionales como herramienta en la enseñanza de la estadística. El tercer capítulo trata sobre algunos aspectos epistemológicos que rigen la inferencia estadística así como su enseñanza. El capítulo cuatro contiene la propuesta para la enseñanza de IC y en el último capítulo se proveen algunas conclusiones y recomendaciones.

1. Evolución del concepto de intervalo de confianza

En los cursos regulares de estadística se enseña que para inferir respecto a una característica de la población, denominada parámetro, se hace uso de su estimación a partir de los datos encontrados en la muestra. Ante la imposibilidad de asegurar que este cálculo sea efectivamente lo acontecido en la población, se prefiere proponer un rango de valores que tengan la posibilidad de contener al parámetro. Esto se logra generalmente mediante un IC que es entendido como un conjunto de *valores* (calculado a partir de los *datos* de una *muestra*) en el cual puede encontrarse el *verdadero valor* del parámetro con un determinado nivel de confianza. El *coeficiente de confianza* indica el *porcentaje de las muestras,* tomadas en las mismas condiciones, en las cuales el intervalo cubriría el verdadero valor del parámetro.

El IC como procedimiento, describe una regla general de construcción de dicho rango de valores, a partir de una *estadística* calculada con los datos de la muestra, para el parámetro correspondiente. La *estadística* tiene una distribución de muestreo conocida que indica la manera en la cual se comporta en el muestreo repetido e involucra conceptos de probabilidad. La idea general de IC se particulariza dependiendo del parámetro a estimar (media, proporción, varianza, etc.) y según el tipo de distribución de donde se obtiene la muestra (discreta o continua) así como condiciones de simetría y otros aspectos que se conoce de la misma.

El IC es uno de los procedimientos que en unión a la prueba de hipótesis, conforman los mecanismos clásicos de la inferencia estadística, entendida esta como el conjunto de aspectos teóricos y de técnicas y métodos que permite emitir conclusiones a nivel de la población usando la información de una muestra. El proceso de estimación de parámetros considera que un determinado fenómeno aleatorio esta caracterizado por una distribución de probabilidad, la cual depende de uno o varios parámetros que se consideran *constantes y desconocidos*; al no ser posible recolectar los datos de toda la población, se utiliza una muestra aleatoria, de esta misma población para hacer la estimación. El problema consiste en proveer un valor aproximado del parámetro (o parámetros) a partir de los datos observados del estadístico (o estadísticos) en la muestra. Es decir, considerando que $X_1, X_2, \cdots, X_n$ es un conjunto de n variables aleatorias cuyos valores particulares pueden ser obtenidos a través de muestras aleatorias y que tienen una función de probabilidad que depende de k parámetros,

$$f(x_{1,} x_2, \cdots, x_n / \theta_{1,} \theta_2, \cdots, \theta_k) \tag{1.1}$$

el objetivo es estimar estos parámetros teniendo en cuenta los valores observados para una realización particular de las n variables aleatorias $x_1^{'}, x_2^{'}, \cdots, x_n^{'}$.

La estimación de un parámetro θ mediante un IC usando el llamado método pivotal (Canavos, 1988, pág. 273), consiste en encontrar una estadística T que utilice y resuma toda la información proveniente de la muestra aleatoria con respecto a θ y disponer de una relación con otra variable $X = f(T; \theta)$, llamada variable aleatoria pivotal, que involucra a θ pero cuya distribución de probabilidad no depende de θ ni de otro parámetro desconocido. De la distribución de X se pueden seleccionar dos valores x_1 y x_2 que cumplan la siguiente condición,

$$P(x_1 < X < x_2) = 1 - \alpha, \tag{1.2}$$

donde $1 - \alpha$ es el coeficiente de confianza. Sustituyendo la variable aleatoria pivotal y con el manejo algebraico adecuado se puede expresar el evento como,

$$P[h_1(T) < \theta < h_2(T)] = 1 - \alpha, \tag{1.3}$$

con $h_1(T)$ y $h_2(T)$ funciones de la estadística T y por consiguiente variables aleatorias. El intervalo de confianza para θ se obtiene sustituyendo en $h_1(T)$ y $h_2(T)$ los estimadores calculados con los datos de la muestra, dando origen al llamado intervalo de confianza bilatetral.

La presentación anterior es la forma como actualmente se trabaja la noción de IC, pero esta recorrió por una serie de transformaciones. A continuación se presenta la manera en la cual evolucionó el concepto de IC iniciado con la visión Bayesiana, luego con la incorporación de la probabilidad fiducial y finalmente con el enfoque frecuentista.

1.1 La visión de Bayes

Uno de los primeros métodos para la solución del problema de la estimación se debe a Thomas Bayes (1702-1761) quien en su teorema (Devore, 2008, pág. 72) especifica la manera en la cual es posible calcular las probabilidades a posteriori de un efecto ya ocurrido, condicionado a una determinada causa. Se toma en cuenta para la solución dos fuentes de información, las probabilidades a priori de los eventos que se consideran causales y las probabilidades que los eventos causales produzcan el efecto de interés.

La contribución de Bayes reside en considerar los parámetros a estimar como variables aleatorias. Este autor plantea una serie de experimentos donde aparece p veces el evento, q veces su complementario y se considera que el parámetro es X, la probabilidad del evento. El objetivo radica en determinar la probabilidad de que X se encuentre entre x_1 y x_2. Para Bayes, la probabilidad en cuestión viene dada por la expresión siguiente:

$$P(x_1 \leq X \leq x_2) = \frac{\int_{x_1}^{x_2} x^p(1-x)^q dx}{\int_0^1 x^p(1-x)^q dx} \quad (1.4)$$

Al considerar que el parámetro es desconocido y estar interesado en determinar la probabilidad de que este se encuentre dentro de un conjunto dado de valores, Bayes propone la construcción del llamado intervalo de credibilidad; este es el punto de partida hacia la construcción de los intervalos de confianza (Olivo, Ortiz, & Batanero, Notas historicas sobre los intervalos de confianza e implicaciones didácticas, 2007).

Simon Laplace (1749-1827) trabaja sobre problemas relacionados con la inferencia y estudia en particular problemas relacionados con la estimación del valor de p en una distribución Binomial, para lo cual considera una urna en donde hay bolas blancas y negras en una relación desconocida. Cuando se extrae una bola y se encuentra que es blanca, se desea establecer la probabilidad de que la relación sea p a q. En esta situación el evento es conocido pero las causas desconocidas.

Refinando el problema se considera que F es un evento ocurrido y que hay n causas desconocidas θ_1, θ_2,…, θ_n. Al considerar que el evento puede ser producido por n causas diferentes, entonces el cociente de las probabilidades de esas causas dado el evento, es igual que el cociente de las probabilidades del evento dadas las causas, esto puede se escrito así:

$$\frac{P(\theta_i/F)}{P(\theta_j/F)} = \frac{P(F/\theta_i)}{P(F/\theta_j)}\,. \quad (1.5)$$

Esto lleva a pensar que la probabilidad para cada una de las causas es igual a la probabilidad del evento dada esa causa, dividida entre la suma de todas las probabilidades de los eventos dadas cada una de las causas (Díaz, 2007) y de acuerdo a la notación establecida es:

$$P\left(\theta_j / F\right) = \frac{P(F/\theta_j)}{\sum_{i=1}^{n} P(F/\theta_i)}\,, \tag{1.6}$$

el cual es reconocido como equivalente al Teorema de Bayes y puede ser orientado para calcular $P(\theta_1, \theta_2, \dots, \theta_n / x`_1, x`_2, \dots, x`_{\mathrm{n}})$, es decir la probabilidad a posteriori para $\theta_1, \theta_2, \dots, \theta_n$, dado un conjunto de eventos o valores indicados $x`_1, x`_2, \dots, x`_{\mathrm{n}}$, en aquellas situaciones donde se conocen las probabilidades a priori $P(\theta_1, \theta_2, \dots, \theta_n)$. Este teorema se puede ajustar para calcular la probabilidad que un determinado parámetro θ_i se encuentre dentro de un par de valores, considerando que $\hat{\theta}_i$ es un estimador de θ_i por ser el más probable y γ un valor fijo positivo. La probabilidad así descrita,

$$P\left(\hat{\theta}_i - \gamma \leq \theta_i \leq \hat{\theta}_i + \gamma / x`_1, x`_2, \dots, x`_{\mathrm{n}}\right) \tag{1.7}$$

corresponde a la exactitud en la estimación, es decir evalúa cuan buena es. El talón de Aquiles de esta propuesta radica fundamentalmente en que la probabilidad a priori es generalmente desconocida pero para salvar este escollo se ha sugerido usar la distribución uniforme (Díaz, 2007).

Tanto Laplace como Gauss contemplan un valor desconocido para el parámetro θ y un cierto número de mediciones x_i todas sujetas a error aleatorio. Fue común en estos autores considerar la función de perdida $L\left(\hat{\theta}, \theta\right)$, que representa la diferencia que se admite por adoptar a $\hat{\theta}$ como estimador de θ. La función de pérdida que uso Laplace fue el valor absoluto de la diferencia, $L_L\left(\hat{\theta}, \theta\right) = \left|\hat{\theta} - \theta\right|$, mientras que la utilizada por Gauss fue la perdida cuadrática $L_G\left(\hat{\theta}, \theta\right) = \left(\hat{\theta} - \theta\right)^2$, de donde se generó la teoría de los mínimos cuadrados muy reconocida particularmente en la estimación del modelo lineal.

1.2 Probabilidad fiducial de Fisher

El uso del teorema de Bayes requiere el conocimiento de probabilidades a priori, esto conlleva el problema de que esta información no está usualmente disponible para el parámetro, por lo cual se acude a definiciones subjetivas de la misma. Fisher en su artículo *Inverse Probability* que data de 1930, introdujo la noción de intervalos fiduciarios. La llamada probabilidad fiducial cuantifica el chance que tiene el parámetro de tomar un determinado valor, enfatizando que el parámetro es una variable. A manera de ejemplo considerar una población normal con σ conocida y el valor $\mu_1 = \bar{X} - 1.64\frac{\sigma}{\sqrt{n}}$; lo que la probabilidad fiducial afirma es que $P(\mu < \mu_1) = 0.95$ y cuyo significado es que en el 5% de las posibles muestras, μ no supera el valor μ_1 obtenido en los datos.

Entonces si se considera una probabilidad P, se puede establecer una relación entre el estadístico y el parámetro, tal que $\hat{\theta}$ es el valor en el P% percentil correspondiente a un θ dado. Tomando una probabilidad de 0.95, implica que en el 5% de las muestras $\hat{\theta}$ excederá al percentil 95% asociado al valor de θ de la población de donde se extrajo la muestra. Esta relación es denominada por Fisher como el “5% fiducial del valor de θ”. Esta teoría de Fisher permite calcular valores de verosimilitud sobre los parámetros, pero no proporciona una distribución de probabilidad respecto a los parámetros desconocidos.

1.3 La replicación de experimentos de Neyman

La teoría moderna sobre los intervalos de confianza comienza con el trabajo de Jerzy Neyman quien en su artículo “On the two different aspects of the representative methods” hace referencia por primera vez a este término (Olivo, Ortiz, & Batanero, Notas historicas sobre los intervalos de confianza e implicaciones didácticas, 2007). La tesis plantea que el problema de estimación puede ser resuelto con base en la teoría frecuentista de probabilidades, sin requerir el uso de probabilidades a priori y la solución consiste en determinar un intervalo dentro de cuyos límites se supone que esta contenido el valor del parámetro. La valides de esta teoría está sustentada en la ley de los grandes números, al suponer que siempre que se repita un experimento con una probabilidad constante *P* de cierto resultado A, a la larga la frecuencia relativa de la aparición de A se aproximará a *P*. Para este procedimiento se asume que la estimación $\hat{\theta}$ no será exactamente igual al parámetro θ, siendo necesaria una medida de la exactitud de la estimación, usando para este fin a $S_{\hat{\theta}}$, la desviación con respecto a $\hat{\theta}$; luego se calculan dos estimaciones de la forma:

$$LI = \hat{\theta} + k_1 S_{\hat{\theta}} \quad \text{y} \quad LS = \hat{\theta} + k_2 S_{\hat{\theta}} \quad , \tag{1.8}$$

donde LI es la estimación inferior, LS es la estimación superior, k_1 y k_2 constantes asociadas a los límites dentro de los cuales se presume que cae el verdadero valor del parámetro. Se necesita elegir por anticipado un valor ε para el coeficiente de confianza ($0<\varepsilon<1$), que en los procedimientos actuales de estimación se acostumbra a designar como $1-\alpha$. Neyman se esfuerza por distinguir que el coeficiente de confianza está referido a los valores de la estimación superior e inferior pero de ninguna manera a la probabilidad de que el valor del parámetro se encuentre dentro de tales limites (Mayo, 1981). En otras palabras, la noción de intervalo de confianza así presentada considera al parámetro constante, mientras que los límites del intervalo son variables que cambian de una muestra a otra.

1.4 Consideraciones generales

En resumen, es útil para los procesos de enseñanza y aprendizaje examinar las posturas de los autores y el contexto histórico que rodean la consolidación del concepto de intervalo de confianza dentro de una teoría de la estimación. En este caso la estimación de parámetros que inicialmente se dio con la construcción de intervalos de credibilidad con una visión bayesiana, posteriormente Fisher promulgó los intervalos fiduciales, y finalmente Neyman planteó el concepto de intervalos de confianza. Entre las posturas acerca de esta noción, se consideró en primera instancia el parámetro como una variable aleatoria, los límites del intervalo de credibilidad fijos y la probabilidad referida al chance de que el parámetro estuviese contenido en el intervalo en un único experimento. Esto se fue transformando hasta considerar el parámetro es fijo, los límites aleatorios y la probabilidad interpretada en términos frecuenciales, indicando la proporción de intervalos que contienen el parámetro en la supuesta repetición indefinida del experimento, mediante la obtención de muestras de la misma población.

2. La comprensión del concepto de intervalo de confianza

En este aparte se presentan algunos resultados publicados respecto a las dificultades que presentan investigadores y estudiantes de nivel universitario al enfrentar el concepto de IC. Algunos de estos resultados son compartidos por las pruebas de hipótesis estadísticas, toda vez que estos son dos procedimientos que están relacionados y tienen un tronco común: las distribuciones de muestreo y la probabilidad. Por otra parte, algunos de los problemas referidos han sido también evidenciados en la práctica docente, lo que resulta ser útil pues se dispone de experiencias que apoyan el estudio de la situación objeto de investigación.

2.1 Dificultades en la comprensión

Las indagaciones sobre la comprensión de IC se han justificado en gran medida, porque esta noción se utiliza ampliamente en los métodos estadísticos que son demandados por investigadores en sus publicaciones, siendo indispensable que la interpretación del IC sea la apropiada; esta noción también está presente en el currículo de estudiantes a nivel de pregrado y posgrado. Tanto para investigadores como para estudiantes hay evidencias de dificultades en la comprensión del IC.

En un publicación que estudió los errores en la interpretación de IC por parte de investigadores (Cumming, Williams, & Fidler, 2004), se empleó una aplicación computacional donde se mostraba a los participantes un diagrama de una media muestral con el correspondiente IC del 95% y se solicitaba ubicar donde se considera más probable que se presentara la media de la muestra si el experimento se replicara 10 veces. Los resultados indicaron que los investigadores tienen una mala concepción del nivel de confianza, manifestada en la creencia errónea de que un IC del $(1-\alpha)100\%$ captura en esta misma proporción los promedios de una serie de muestras aleatorias.

Los investigadores creen que un intervalo de confianza del 95% incluirá alrededor del 95% de las medias muestrales en las futuras replicaciones, el verdadero valor es 83.4% (Estes, 1997). Según lo anterior, no el 5% como lo creen los investigadores, es en realidad el 16.6% de las medias de las replicaciones las que quedan por fuera de un intervalo de confianza del 95%. En resumen los investigadores esperan (erróneamente) una **alta probabilidad de replicación**, suponiendo que, al tomar una nueva muestra, el nuevo intervalo de confianza será **muy parecido al original**.

En otra investigación (Cumming & Fidler, 2005) esta vez con estudiantes universitarios, se solicitó a 55 alumnos del programa de ciencias del ambiente interpretar resultados publicados en revistas científicas referidos a IC y PH. Concerniente a las pruebas de hipótesis se analizaba específicamente el valor *p* entendido así: "Si W es un estadístico de prueba, el valor *p* o nivel de significancia alcanzado, es el nivel mínimo de significancia α para el cual los datos observados indican que se debe rechazar la hipótesis nula" (Wackerly, Mendenhall, & Scheaffer, 2002)

Entre los resultados que obtienen los autores se destaca que 44% de los estudiantes interpreta incorrectamente un valor *p* pequeño en un contraste, al favorecer en la conclusión lo consignado en la hipótesis nula. Cuando se presentaron los resultados que los estudiantes deberían analizar pero esta vez usando IC, sólo 18% interpreta incorrectamente los resultados. Esto puede sugerir que la comprensión de los intervalos de confianza podría ser más accesible que interpretar un contraste de hipótesis. En el mismo estudio, se encontró que aquellos estudiantes que se enfrentaron al análisis de los IC y posteriormente a los del valor *p*, tienden a responder acertadamente con una mayor frecuencia que aquellos que observaron en primera instancia el valor *p*.

Otro objetivo en esta investigación trataba sobre la manera en la cual se concibe la naturaleza de los intervalos de confianza, para lo cual se encuestó a 180 estudiantes de psicología quienes habían tomado por un año un curso de introducción a la estadística enfocado en intervalos de confianza. Una proporción importante consideran los intervalos solo como estadísticos descriptivos, ignorando su naturaleza inferencial. Prueba de ello fue que 38% de ellos creían que los IC proveen valores posibles para la media de la muestra; por otra parte el 19% asociaban el intervalo con rango o un rango truncado para los valores de la variable. Se presentan también dificultades en la manera en la que se relacionan entre sí los distintos conceptos que intervienen en los intervalos de confianza. El 29% de los estudiantes considerados creen que el ancho del intervalo de confianza se incrementa al aumentar el tamaño de muestra, un 20% manifiesta que no se altera, el 36% desconoce la relación entre ancho del IC y el tamaño de muestra, lo que finalmente deja solamente un 15% que refiere acertadamente la relación entre el ancho del IC y el tamaño de muestra.

Estos resultados, además brindan la posibilidad de aportar a la comprensión de este concepto en los estudiantes. Las observaciones subjetivas propias y las reportadas en los estudios refuerzan la idea que el aprendizaje de IC amerita ser estudiado a profundidad. El concepto y su apropiación por parte de los estudiantes han generado dificultad. Es un tema que incluye gran diversidad de conceptos estadísticos previos finamente entrelazados así como otros propios de la inferencia.

2.2 Uso de software y simulación como alternativa

Si bien la simulación es un recurso didáctico usado para ayudar a comprender la noción de IC, de ninguna manera puede proveer justificaciones o demostraciones. Uno de los proyectos pioneros al respecto fue “Exploring the role of computer simulations in developing understanding of sampling distributions (delMas, Garfield, & Chance, 1999) en el cual se analiza el uso de la simulación mediante el software Sampling SIM (http://www.tc.umn.edu/~delma001/stat_tools/software.htm) para inferencia y muestreo. Primero el estudiante debía responder un cuestionario de entrada para determinar su nivel de conocimiento previo. Posteriormente se planeó una serie de actividades en donde mediante simulación se conducía al estudiante por diversas escenarios con el objetivo de comprender los intervalos de confianza, en última etapa se aplicó una prueba final que involucra confrontación con algunas de las preguntas de la prueba de entrada, temas relacionados con las actividades de simulación y sobre problemas de aplicación. Se usó la experiencia anterior para hacer mejoras al software (delMas, Garfield, & Chance, 2004), se rediseño la prueba de entrada y se mejoraron las actividades buscando que los estudiantes que hicieran sus propias predicciones y comparaciones, las cuales se evaluaron mediante este mismo software.

Otro software para el aprendizaje de estadística inferencial es HyperStat (http://davidmlane/hyperstat/), en cuyo vinculo el profesor David Lane, adscrito a la Universidad de Rice en Houston Texas, proporciona un conjunto de aplicaciones para el aprendizaje de conceptos estadísticos básicos, entre los que se encuentra un simulador para generar intervalos de confianza provenientes de una distribución normal tomando muestras de tamaño 10, 15 y 20.

La Universidad Complutense de Madrid en la facultad de biología, tiene una aplicación en línea (http://e-stadistica.bio.ucm.es/) que apoya la comprensión de la estadística, y particularmente en lo referente a los intervalos de confianza diseñó un aplicativo que genera muestras aleatorias de una distribución normal estándar. En Cumming (2012) se anexa un software llamado ESCI, que trabaja sobre Microsoft Excel y que está diseñado para realizar simulaciones y apoyar la comprensión de de intervalos de confianza, pruebas de hipótesis, muestreo y metanálisis.

Desde la década de los 80 y con más fuerza desde los 90´s se ha intensificado el uso de software como apoyo para la docencia. La enseñanza moderna de estadística y análisis de datos supone el manejo de algún tipo de software estadístico, existiendo actualmente una amplia variedad de opciones para este fin. Para un análisis introductorio acerca de las posibilidades de empleo de las herramientas informáticas en el contexto de la enseñanza, se comienza por señalar algunas distinciones básicas relacionadas con las características del software. Una primera cuestión tiene que ver con el alcance del software en cuanto al rango o variedad de métodos estadísticos que manejan. En una categoría esta los programas de múltiple propósito como:

- SPSS® (http://www-01.ibm.com/software/analytics/spss/),
- SAS® (http://www.sas.com/),
- STATA® (http://www.stata.com/).

Otro grupo lo conforman los programas especializados en la aplicación de una técnica estadística, como por ejemplo:

- Lisrel (http://www.ssicentral.com/lisrel/index.html) para modelado de ecuaciones estructurales y manejo de datos de encuestas con muestras complejas.
- AMOS (http://www-142.ibm.com/software/products/es/es/spss-amos/) para modelos de ecuaciones estructurales
- HLM (http://www.ssicentral.com/hlm/) para modelos lineales jerárquicos,

También es conveniente distinguir entre el software estadístico orientado principalmente a realizar aplicaciones a nivel de usuario y los lenguajes de programación estadística, más apropiados para la innovación y el desarrollo de métodos en estadística computacional, dos exponentes de este tipo son:

- R (http://www.r-project.org/)
- LispStat (http://www.divms.uiowa.edu/~luke/xls/xlsinfo/)

Estos dos últimos son poderosos lenguajes de programación estadística, que requieren un cierto grado de experticia, más propio de un usuario avanzado que de un estudiante, sin embargo R se encuentra en plena expansión y se espera que en el futuro próximo avancen sensiblemente los desarrollos en materia de su interfaz gráfica de usuario.

Otro elemento clave y determinante para la elección de un software es su accesibilidad, es decir, las posibilidades de ser adquirido y las condiciones bajo las cuales se puede distribuir. Cabe mencionar que algunos paquetes estadísticos profesionales son prácticamente inaccesibles para docentes y estudiantes en nuestro ambito, por lo anterior los softwares libre como R, son ahora muy apreciados.

Si bien las restricciones acerca del uso de programas licenciados puede ser una limitante, es también importante examinar otras características que son relevantes al momento de elegir un software estadístico. Salas (2008) hace un estudio comparativo entre SPSS(http://www-01.ibm.com/software/analytics/spss/),, SAS(http://www.sas.com/), y Rhttp://www.r-project.org/) sobre los siguientes ocho aspectos: amigabilidad con el usuario, manipulación de datos, calidad de gráficos, control de procesos, costo, variedad de análisis estadísticos, documentación y soporte de ayuda, y sistemas operativos. El estudio deja muy bien librado a R (http://www.r-project.org/) y entre las conclusiones se resalta que "*la gran versatilidad de los procedimientos estadísticos disponibles (así como los tarea-específicos), la capacidad de producir gráficos de calidad y la amplia documentación gratuita, entre otros aspectos, hacen de R un excelente programa estadístico para ser usado en docencia e investigación*" (Salas, 2008, pág. 230).

En este estudio se decidió usar R como herramienta para la enseñanza del concepto de IC en particular se usa simulación como método para estudiar el impacto que sobre los intervalos de confianza tienen el tamaño de muestra y distribución de donde provienen los datos. En publicaciones sobre didáctica de la estadística, la simulación ha ganado un espacio importante, esto motiva la realización del trabajo. Aunque se reconoce que el libro de texto es aún el principal recurso usado en el aula de clase, se puede abordar el estudio de los IC mediante la simulación, cuyo uso es intuitivamente deseable e incluso aparece como una opción a explorar en los contenidos programáticos de las asignaturas de estadística.

La aparición en las aulas de las herramientas computacionales ha permitido, por un lado, que tiempo de clase sea aprovechado mejor y por el otro, que el estudiante tome parte activa en el proceso de enseñanza-aprendizaje, Al respecto se considera que “la informática puede jugar un papel importante, no para reemplazar al profesor, pero si para brindarle apoyo en su tarea y darle una opción más al estudiante para la consolidación de su aprendizaje” (Behar & Grima, 2001). Este mismo autor menciona que el desarrollo de los contenidos debe ir acompañado de estrategias que fortalezcan el entendimiento de los conceptos y reforzados en la simulación como apoyo para “ilustrar de una manera más vivencial el significado de la teoría”.

3. Aspectos epistemológicos

En este aparte se puntualizan algunos aspectos medulares que rigen la lógica de la inferencia estadística, para esto se esbozan brevemente los principios que sirven de base para el proceso de inferencia (Gutiérrez Cabria, 1994).

El Principio de Aleatoriedad admite dos versiones: (i) todo individuo de la población tiene igual probabilidad de pertenecer a la muestra, (ii) muestras de igual tamaño tienen la misma probabilidad de aparecer en el muestreo de una población. De las nociones básicas de probabilidad se puede deducir que cuando hay independencia (muestreo con reemplazamiento), las versiones coinciden. Puesto que en muchas investigaciones el espacio de muestreo puede constar de diversas variables aleatorias, entonces al tomar una muestra aleatoria, esta resulta distinta de lo que ocurre cuando la muestra se toma del espacio de valores que está formado por todas las variables, por lo cual es de interés reducir el espacio de muestreo, así los principios que se mencionan a continuación permiten llevar a cabo esta reducción, esto son: el principio de suficiencia, el principio de condicionalidad y el principio de verosimilitud.

La lógica de la inferencia estadística está interesada el estudio de los principios que gobiernan el problema consistente en realizar una apropiada conclusión a partir de los datos disponibles. Su principal interés es determinar procedimientos y reglas para llevar a cabo la inferencia; una de tales reglas formula un enunciado de cierta naturaleza (por ejemplo, una estimación puntual, o que el parámetro θ pertenece a un cierto conjunto, o la decisión de tomar cierta acción) sobre cada valor concebible de la variable observada X; este enunciado es la inferencia nominal. Luego cuando se aplica la regla especificada a unos datos, a la inferencia nominal se le adjunta información sobre el comportamiento de la regla, el cual viene medido en términos de probabilidad. Esta información complementaria se denomina inferencia estocástica.

En la estimación por intervalo la inferencia nominal es una aseveración que dice que θ pertenece a cierto conjunto dependiente de los datos, por ejemplo, al conjunto $\hat{\theta}(X_1, \cdots, X_n) = [\bar{x} - a, \bar{x} + a]$. A esta inferencia nominal es preciso añadir que la probabilidad de que el intervalo comprenda a θ adquiere cierto valor llamado coeficiente de confianza; esta segunda aseveración constituye la inferencia estocástica. La practica actual de la inferencia estadística implica apelar a alguna regla de inferencia, establecer las propiedades de las distribuciones probabilísticas y generar las inferencias nominal y estocástica, implicadas por los datos disponibles.

Desde el punto de vista teórico y práctico, los estadísticos desarrollan métodos y técnicas, desde posiciones ideológicas distintas, pero los principios pueden aportan para tener una base filosófica común. Se tienen objetivos e instrumentos comunes, pero las preguntas de los estadísticos pasan por establecer las calidades de la muestra, e incluso indagarse ¿Qué es una muestra aleatoria? ¿Apoyan los datos al modelo propuesto? ¿Qué parámetros son explicables condicionados a los datos disponibles?, pero las posiciones al respecto provienen de escuelas distintas.

Según Gutiérrez (1994) los estadísticos plantean el problema de la inducción en forma distinta a la expuesta por los filósofos. Un problema clásico de inducción en la estadística plantea que si se han extraído balotas de una urna, previa aleatorización; ¿Cuál es la probabilidad de que salga blanca en la n-ésima extracción, dado que ha salido blanca en las (n-1) extracciones anteriores?. Para la inducción, los estadísticos deben disponer de una población especificada y finita, condición necesaria para la aleatorización. Aunque esta restricción es muy rigurosa, se han propuesto esquemas de muestreo donde cada unidad elegida tenga asociada una probabilidad, que no necesariamente sea la misma para todas.

Por otra parte, se requiere también establecer ¿cuál es la tendencia que ha servido de guía a la propuesta de enseñanza?. Para comenzar, se sabe que en la investigación en didáctica de las ciencias se ha identificado una serie de dificultades ya clásicas en el proceso de enseñanza: la estructura lógica de los contenidos conceptuales y su exigencia formal, la influencia de los conocimientos previos y las preconcepciones del alumno. Esta visión ha cambiado para centrarse en las concepciones epistemológicas de los alumnos, sus estrategias de razonamiento y la metacognición.

Las concepciones epistemológicas tienen que ver como se estructura, evoluciona y produce el conocimiento; en el mundo de los estudiantes gira para concebir las concepciones sobre cómo se aprende el conocimiento científico (Campanario & Moya, 1999). De acuerdo con lo anterior se evidencian concepciones en los alumnos que los llevan a pensar la ciencia articulada en términos de ecuaciones, definiciones y teoremas que deben ser memorizados más que comprendidos.

También hay evidencia que las estrategias de análisis de los problemas son superficiales, con una actitud poco critica lo cual se hace latente cuando los profesores ya no se sorprenden de errores de observación en sus alumnos o ante preguntas que requieren respuestas bien elaboradas y que demanda tiempo, se contestan en intervalos sorprendentemente cortos (Campanario & Moya, 1999). Este mismo autor, hace ver que ante las diversas posibilidades que se disponen, cabe la pregunta de ¿cómo enseñar?. Este resulta ser un problema abierto y complejo. En este documento se opta por el diseño de unidad didáctica, en la cual confluye tanto los aspectos teóricos como los que llevan a las acciones. El diseño involucra una serie de componentes, análisis científico, análisis didáctico, objetivos, estrategias didácticas y estrategias de evaluación.

En el análisis científico se hace la selección de los contenidos, procedimientos científicos y las actitudes. Para el análisis didáctico se recurre a la indagación de las concepciones previas de los alumnos, examen de las exigencias cognitivas de los contenidos, la delimitación de las aplicaciones para la enseñanza. Sobre las estrategias didácticas se prefiere diseñar una secuencia general de enseñanza, protocolizar las actividades de enseñanza y la elaboración de material de aprendizaje.

4. Propuesta de Enseñanza

La propuesta que se describe a continuación está compuesta por actividades que apoyan la construcción del concepto de intervalo de confianza. Se tiene como característica principal el uso de la simulación, que aquí se utilizó para extraer un conjunto de muestras aleatorias provenientes de una distribución conocida y a partir de cada muestra construir un intervalo de confianza.

La primera actividad considera que la distribución de la cual se extraerán las muestras aleatorias es Binomial. La segunda actividad está referida a la generación de intervalos para una distribución normal. La principal razón para elegir estas dos distribuciones es estimar el promedio tanto para una distribución discreta como para una continua.

4.1 Intervalos de confianza para una distribución Binomial

Si un experimento aleatorio consiste de r ensayos repetidos tales que: (1) los ensayos son independientes, (2) cada ensayo tiene solo dos resultados posibles, denominados "éxito" y "fracaso", (3) la probabilidad de "éxito" es p y la de "fracaso" es $1-p$, (4) la probabilidad de "éxito" permanece constante a lo largo de los r ensayos y (5) se considera la variable X como el número de éxitos en los r ensayos: Entonces X tiene distribución Binomial con función de probabilidad,

$$f(x) = \binom{r}{x} p^x (1-p)^{r-x}, \quad x = 0,1,2,\cdots,r \quad , \tag{4.1}$$

donde $\binom{r}{x} = \frac{r!}{(r-x)!x!}$ y los valores de r y p son los parámetros de la distribución. La notación para esta distribución es $X \sim Bin(r, p)$ y su promedio y varianza son, respectivamente:

$$\mu = r * p \qquad \sigma^2 = r * p * (1 - p)\ . \tag{4.2}$$

Para generar las simulaciones se consideró X como el número de éxitos que se observan en r=8[1] ensayos, cada realización de la variable aleatoria genera un valor de cero hasta ocho. Se tomaron tres valores para p, a saber: 0,1, 0,5 y 0.9. Se quiere estimar el promedio de la distribución Binomial (ver ecuación 4.2) usando intervalos de confianza, cuando se replica el experimento en 8 ocasiones. Para estimar el promedio de la distribución binomial mediante un IC, se usa la expresión siguiente (Wackerly, Mendenhall, & Scheaffer, 2002, pág. 386):

$$\bar{x} \pm Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad , \tag{4.3}$$

donde, $\bar{x}$ es el promedio de éxitos en n replicaciones del experimento, σ es la desviación estándar obtenida a partir de (4.2) y $z_{1-\alpha/2}$ es el cuantil de una distribución normal estándar, que satisface:

$$P\left(Z < z_{1-\alpha/2}\right) = 1 - \alpha/2 \tag{4.4}$$

Los IC se obtienen mediante simulación de muestras aleatorias de tamaño diferente (n=10, 50 y 90) y proveniente de una distribución Binomial (8, p).

POBLACIÓN OBJETIVO

Estudiantes universitarios inscritos en el curso de bioestadística donde se cobijan temas básicos de estadística y probabilidad así como elementos de inferencia. En este curso se matriculan estudiantes de las carreras de Biología Marina y Biología Ambiental. Al momento de tomar el curso los estudiantes tiene alrededor de cuatro semestres de permanencia, y con los elementos básicos de la matemática dado por los cursos de precálculo y cálculo.

PRÁCTICA DE AULA

Definición del experimento Binomial: Si un investigador se interesa en el siguiente experimento: siembra 8 semillas y desea considerar la siguiente variable aleatoria X:número de semillas que germinan cuando se siembran 8 de ellas. Entonces los resultados posibles pueden variar desde la situación menos favorable para el

[1] El valor usado para r se tomó a discreción y solo para efecto de cálculo.

investigador, cuando ninguna germina (x=0), hasta la más deseable, en donde todas germinaron (x=8).

Modo de trabajo: Se solicita a los educandos conformen grupos de máximo tres estudiantes; grupos de mayor número dificulta una participación activa en la práctica. Después de cada actividad, el profesor efectúa una breve revisión de los cálculos solicitados para minimizar los errores de procedimiento.

Actividad 1. (Tiempo disponible 10 minutos)

Describir en términos prácticos cual es el impacto sobre la cantidad de semillas que germinan, si la probabilidad de que cada una de ellas brote toma uno de los siguientes valores: 0,1, 0,5, y 0,9. Particularmente responda la pregunta siguiente: ¿Cual es número promedio de semillas que germinan y la desviación estándar para cada uno de los tres casos? Consigne los resultados y el análisis respectivo en la tabla siguiente:

***p*=0,1**	***p*=0,5**	***p*=0,9**

Objetivos de aprendizaje: esta actividad busca explorar los conceptos previos de los estudiantes y a partir del cálculo del promedio de semillas que germinan según sea la probabilidad que tenga cada una de brotar. Examinar si lograr intuir que hay una relación directa entre la cantidad de semillas que germinan y la probabilidad de que cada una brote.

Actividad 2. (Tiempo disponible 10 minutos).

Suponer que el experimento se ha repetido 10 veces, es decir se tomó 10 grupos cada uno de 8 semillas y se tienen los resultados mostrados en la primera fila. Los datos se logran mediante la función rbinom(10,8,p) de R con p=0.1, 0.5 ó 0.9. Escriba para cada caso en la fila 2, ¿cuál es el valor de p que más se ajusta a los resultados mostrados? Explique su respuesta en la fila 3.

5, 2, 5, 6, 4, 4, 5, 3, 2, 3	2, 1, 0, 1, 2, 1, 1, 3, 1, 2	6, 8, 7, 8, 8, 8, 8, 8, 7, 8
***p*=**	***p*=**	***p*=**

Objetivos de aprendizaje: Esta actividad tiene particular interés en determinar si a partir de un conjunto de 10 replicas del experimento, los estudiantes logran deducir apropiadamente la probabilidad con la cual se han obtenido los resultados.

Actividad 3. (Tiempo disponible 10 minutos)

Usando la información anterior y para cada caso, estime el promedio de semillas que germinaron mediante un intervalo de confianza del 95%; use z=1.96 y la desviación estándar de cada distribución (detalle los pasos en la fila 2 de la tabla). Adjunte además la interpretación respectiva en la tercera fila.

p=0,1	**p=0,5**	**p=0,9**

Objetivos de aprendizaje: Se pretende que los estudiantes sean capaces de usar la expresión para calcular el IC y que proporcionen una adecuada interpretación dentro el contexto del problema.

Actividad 4. (Tiempo disponible 10 minutos)

Use la función rbinom de R para obtener una segunda muestra de tamaño 10 para cada uno de los valores de p considerados (consigne los valores obtenidos para la muestra en la fila 2) y vuelva a calcular el intervalo solicitado en el paso anterior, usando para ello la fila 3 de la tabla.

p=0,1	**p=0,5**	**p=0,9**

Objetivos de aprendizaje: Esta actividad tiene como finalidad familiarizarse con el manejo de R para generar muestras aleatorias y experimentar la gran agilidad de este software. Por otro lado se pretende concebir un escenario para la aparición de inquietudes teniendo como base las diferencias que puedan observar los estudiantes en los intervalos logrados, comparados con los calculados en la actividad anterior.

Actividad 5. (Tiempo disponible 10 minutos)

Discuta las siguientes preguntas referidas a las dos muestras que se lograron anteriormente:

¿Se encuentra el verdadero promedio contenido dentro de cada uno de los dos intervalos de confianza calculados? ¿En caso contrario a que puede atribuirse? ¿Con que confianza debería capturar el IC el real valor del promedio?

Compare el promedio de la muestra proporcionada y la que se propuso hallar en el paso 4? ¿Deben dar iguales? ¿En el caso de no serlo, es indicativo de que algo anda mal con la muestra y debería desecharse?

Objetivos de aprendizaje: en este momento los estudiantes deben ser capaces de reconocer que no necesariamente el IC tiene que cobijar al parámetro y en el caso de suceder este hecho puede proporcionarse una explicación sobre una base probabilística. Además se invita a la reflexión sobre la creencia, por demás errónea, que al replicar la muestra el IC debería ser similar al obtenido previamente.

Actividad 6. (Tiempo disponible 60 minutos)

A continuación se presenta una grafica con los resultados obtenidos mediante simulación de la distribución Binomial X, número de semillas que germinan al sembrar 8 de ellas. Allí se presentan 100 intervalos de confianza variando el tamaño de muestra según tres posibilidades (10, 50 y 90) y la probabilidad de germinar según 3 opciones (0,1, 0,5 y 0,9). Se tienen 9 combinaciones posibles según varía el tamaño de muestra y la probabilidad de germinar; así los escenarios posibles inician con tamaño de muestra de 10 y p=0,1 hasta tamaño de muestra de 90 y p=0.9. Cuando un intervalo de confianza no contiene el verdadero promedio se ilustra con color rojo.

Alrededor de la simulación de estas nueve situaciones se propone resolver por escrito las siguientes preguntas:

- Note que cada una de las nueve graficas que aparecen no tienen estipulado el valor de p ni el tamaño de muestra con el que se obtuvieron los intervalos. Si se requiere asignar estos dos valores a cada una de las graficas que se muestran,

¿mediante que argumento puede saber cual combinación de tamaño de muestra y probabilidad de germinar es la apropiada para cada una de las nueve graficas?

- Determine mediante observación, cuantos intervalos no contienen el verdadero valor para cada una de las nueve combinaciones posibles de tamaño de muestra y valor de probabilidad. Recuerde que los intervalos que no contiene el real valor del parámetro se encuentran ilustrados con color rojo.
- ¿Esperaría usted que la cantidad de IC que no contienen al verdadero valor del promedio de la distribución sea el mismo para cada uno de los nueve casos?
- De la teoría se sabe que si el tamaño de muestra se incrementa, tiende a ser más angosto el intervalo de confianza. ¿cómo se ve esto reflejado en las graficas?

Objetivos de aprendizaje: en esta actividad se pretende que los estudiantes estén en capacidad de reconocer el impacto que tiene la probabilidad que se presente el evento y el tamaño de muestra, sobre los IC para estimar el promedio de la distribución Binomial.

En la teoría se establece que cuando se construye un IC del 95% se espera que al replicar la muestra, un 5% de los IC no cobijen el parámetro; con el ánimo de confrontar las apreciaciones teóricas con las evidencias proporcionadas por la simulación, se indaga a los estudiantes acerca de la cantidad de intervalos que no contiene el parámetro y los argumento que pueden exponerse cuando los resultados no sean idénticos para cada los nueve escenarios que se proponen.

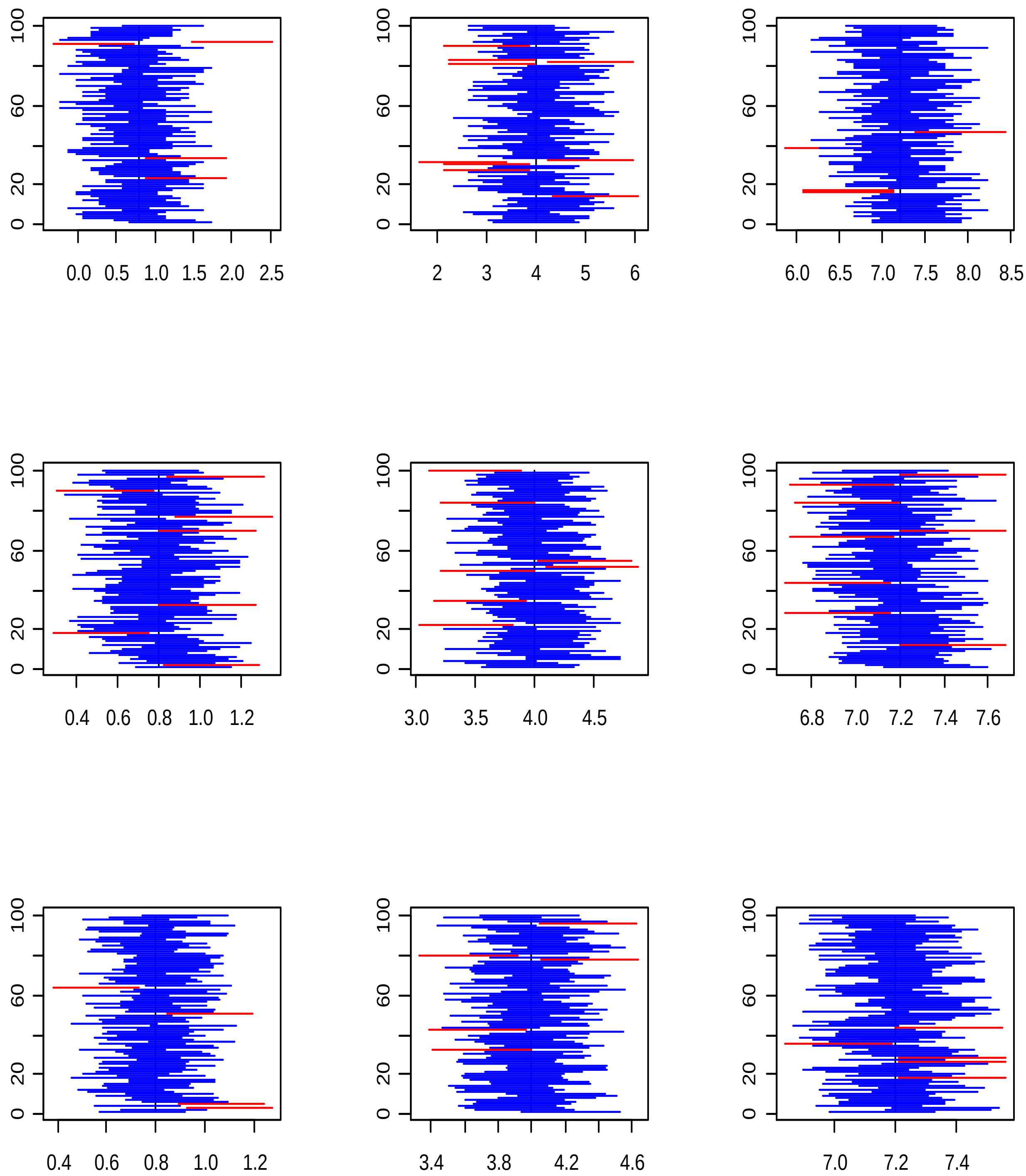

Figura 4.1. Simulación de IC del 95% para X~Bin (8; p), tamaños de muestra 10, 50 y 90 y p=0,1, 0,5 y 0,9.

ESTRATEGIA DIDÁCTICA

Esta práctica requiere que el docente se apoye en una presentación donde se exponga en detalle cada actividad y presenten para cada una las implicaciones, las soluciones analíticas y las conclusiones. Esta propuesta complementa las discusiones, que debe ocurrir con anterioridad, sobre el tema de intervalos de confianza para el promedio.

PRESENTACIÓN ESCRITA

Para apoyar el trabajo de los estudiantes se proveerá una guía de trabajo en donde se consignen los resultados de cada grupo. Esta guía debe incorporar la plantilla en donde cada grupo de estudiantes entregue sus respuestas.

EVALUACIÓN

Una de las características al trabajar con simulación para obtener valores de variables aleatorias, es la diferencia de resultados entre los estudiantes. Entonces el proceso de evaluación se lleva a cabo en dos instantes: discusión global de los procedimientos comunes involucrados y examen de las situaciones que pueden dar origen a variaciones en la respuesta. Por otra parte, es usual que cuando se desea asignar nota a una actividad como la que se propone aquí, se requiere valorar dos aspectos. El primero de ellos, es el examen de los procedimientos y cálculos que se necesitan, y por otra parte las interpretaciones a lugar. En este caso, donde se requieren cálculos previos a las interpretaciones se siguiere examinar los resultados obtenidos; por lo tanto es necesario revisar junto con los estudiantes, que los IC calculados sean los correctos para que las interpretaciones y conclusiones no se vean afectadas por resultados desafortunados.

Según lo anterior, las diferencias entre los estudiantes se verán reflejadas en la calidad de los argumentos y explicaciones solicitadas y en esa medida se asignará la nota respectiva al trabajo de los estudiantes.

4.2 Intervalos de confianza para una distribución Normal

Para este caso se parte de una distribución normal con función de densidad,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; -\infty < x < \infty, \; -\infty < \mu < \infty, \quad \sigma > 0 \tag{4.5}$$

Esta función tiene dos parámetros μ y σ. En esta situación, el objetivo es estimar el promedio de la distribución normal mediante un IC, según la expresión siguiente (Wackerly, Mendenhall, & Scheaffer, 2002, pág. 400):

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \tag{4.6}$$

Donde $t_{1-\alpha/2}$ es el cuantil de la distribución t-student, con n-1 grados de libertad, que satisface:

$$P\left(T < t_{1-\alpha/2}\right) = 1 - \alpha/2 \tag{4.7}$$

Los IC se obtendrán mediante simulación de muestras aleatorias de tamaño diferente y proveniente de distribución normal con un promedio fijo, y distintos valores para la desviación estándar.

PRACTICA DE AULA

Definición del contexto: Se sabe que la envergadura de alas de un ave adulta sigue una distribución normal, con lo cual se espera que una gran proporción de ellas tengan envergaduras de alas cercanas al promedio general y menor proporción de aves con envergaduras alejadas del mencionado valor. Para efecto de describir las características de las estimaciones mediante IC, consideremos que la envergadura promedio es 100 cm.

Modo de trabajo: Se solicita a los estudiantes conformen grupos de máximo tres estudiantes; grupos de mayor número impide una participación activa en la práctica. Después de cada actividad, el profesor efectúa una breve revisión de los cálculos solicitados para minimizar los errores de procedimiento.

Actividad 1. (Tiempo disponible 10 minutos)
Describir en términos prácticos cuál es el impacto sobre las mediciones de la envergadura si la desviación estándar cambia, en particular si toma los siguientes valores: 5cm, 10cm y 15cm. (Ayuda: si una variable tiene distribución normal recuerde

que se puede saber rápidamente la proporción de las mediciones de las envergaduras que están alejadas una, dos y tres desviaciones estándar del promedio)

σ=5	σ=10	σ=15

Objetivos de aprendizaje: esta actividad busca determinar si los estudiantes tiene presente el impacto de la desviación estándar sobre las observaciones de una muestra aleatoria proveniente de una distribución normal. En este sentido se propone la ayuda que muy apropiadamente se suele llamar regla empírica.

Actividad 2. (Tiempo disponible 10 minutos)

Suponga que se ha tomado una muestra aleatoria de 10 aves adultas cuyos resultados aparece en la primera fila de la tabla mostrada a continuación (estas tres muestras fueron simuladas mediante la función rnorm de R con promedio 100 y desviación estándar 5, 10 ó 15).

Escriba para cada caso en la segunda fila, ¿qué valor de la desviación estándar (entre tres posibles valores: 5, 10 o 15) se ajusta más a los resultados mostrados? En la tercera fila explique la razón para la asignación de la desviación estándar que usted propone para cada conjunto de observaciones.

89,3 91,2 112,6 95,8 115,5 138,6 66,8 115,6 128,3 98,2	102,1 105,8 110,3 96,9 100,7 91,4 88,8 95,4 115,9 100,5	102,6 98,7 108,2 95,3 102.7 98,5 101,9 101,3 95,5 104,3
σ=	**σ=**	**σ=**

Objetivos de aprendizaje: como continuación de la primera actividad, se solicita determinar cuál valor de desviación es el apropiado para cada conjunto de datos. Los estudiantes pueden optar por examinar las diferencias entre los datos de cada conjunto, o asignar σ según sea el valor de la desviación estándar muestral.

Actividad 3. (Tiempo disponible 10 minutos)

Usando la información anterior y para cada uno de los tres conjunto de observaciones, estime la envergadura promedio de alas mediante un intervalo de confianza del 95%; use el valor del cuantil de la distribución t-student y la desviación estándar de cada distribución de donde se obtuvo la muestra. Detalle los cálculos requeridos y el IC en la segunda fila y la interpretación respectiva en la fila 3 de la tabla siguiente.

σ=5	**σ=10**	**σ=15**

Objetivos de aprendizaje: Se pretende que los estudiantes sean capaces de usar la expresión para calcular el IC y que proporcionen una adecuada interpretación dentro el contexto del problema.

Actividad 4. (Tiempo disponible 10 minutos)

Use la función rnorm de R para simular una segunda muestra de tamaño 10, con promedio 100 y para cada uno de los valores de σ considerados (escriba la muestra simulada en la segunda fila) y vuelva a calcular el intervalo solicitado en el paso anterior, consignando los resultados en la fila 3 de la tabla.

σ=5	**σ=10**	**σ=15**

Objetivos de aprendizaje: Esta actividad tiene como finalidad familiarizarse con el manejo de R para generar muestras aleatorias y experimentar la gran agilidad de este software. Por otro lado se pretende concebir un escenario para la aparición de inquietudes teniendo como base las diferencias que puedan observarse en los intervalos logrados, comparados con los calculados en la actividad anterior.

Actividad 5. (Tiempo disponible 10 minutos)

Discuta las siguientes preguntas referidas a las dos muestras que se lograron anteriormente:

¿Se encuentra el verdadero promedio contenido dentro de cada uno de los dos intervalos de confianza calculados?. ¿En caso contrario a qué puede atribuirse? ¿Con que confianza debería capturar el IC el real valor del promedio?

Compare el promedio de la muestra proporcionada y la que se propuso hallar en el paso 4? ¿Deben dar iguales? ¿En el caso de no serlo, es indicativo de que algo anda mal con la muestra y debería desecharse?

Si se fija el tamaño de muestra, ¿Para qué valor de la desviación estándar considera que debería ser menor la diferencia entre la media de la muestra y la verdadera media de la población?

Si se fija el valor de la desviación estándar, ¿Para qué valor del tamaño de muestra considera que debería ser menor la diferencia entre la media muestral y la verdadera media de la población?

Objetivos de aprendizaje: con esta actividad se pretende que los estudiantes reconozcan de que no necesariamente el IC tiene que cobijar al parámetro y cuando esto ocurre puede proporcionarse una explicación sobre una base probabilística. Además se invita a la reflexión sobre la creencia, por demás errónea, que al replicar la muestra el IC debería ser similar al obtenido previamente.

Por otro lado se invita a dar explicación sobre el impacto de la desviación estándar y el tamaño de muestra sobre el ancho de los IC.

Actividad 6. (Tiempo disponible 60 minutos)

A continuación se presenta una grafica con los resultados de una simulación para una distribución normal. Allí se presentan 100 intervalos de confianza variando el tamaño de muestra según tres posibilidades (10, 30 y 50) y la desviación estándar según 3 opciones (5, 10 y 15). Así finalmente se tienen 9 combinaciones según varía el tamaño de muestra y la desviación, así los escenarios posibles son: $n=10$ y $\sigma=5$ hasta $n=50$ y $\sigma=15$. Cuando un intervalo de confianza no contiene el verdadero promedio se ilustra con color rojo.

Alrededor de esta simulación se desea resolver por escrito a las siguientes preguntas:

- Determine mediante observación, cuantos intervalos aproximadamente no contienen el verdadero valor del promedio en cada una de las simulaciones? Coincide con lo que se espera si la confiabilidad es del 95%?
- ¿Se espera que la cantidad de IC que no contienen al verdadero valor poblacional sea el mismo para cada uno de los nueve casos?
- Si observa únicamente la primera fila de las simulaciones, explique ¿Cuál es el impacto de la desviación estándar sobre los intervalos mostrados en la grafica? Se aplica también para la segunda fila y tercera fila de simulaciones?
- Si observa únicamente la primera columna de las simulaciones, explique ¿Cuál es el impacto del tamaño de muestra sobre los intervalos hallados? Se aplica también para la segunda columna y tercera columna de simulaciones?

Objetivos de aprendizaje:

En la teoría se establece que cuando se construye un IC del 95% se espera que al replicar la muestra, un 5% de los IC no cobijen al parámetro; con el ánimo de confrontar las apreciaciones teóricas con las evidencias proporcionadas por la simulación, se indaga a los estudiantes acerca de la cantidad de intervalos que no contiene el parámetro y los argumento que pueden exponerse cuando los resultados no son idénticos para cada los nueve escenarios que se proponen.

En las demás preguntas se desea poner de manifiesto, pero en este caso mediante el uso de las graficas, el impacto que tiene la desviación estándar y el tamaño de muestra, sobre los IC para estimar el promedio de la distribución normal

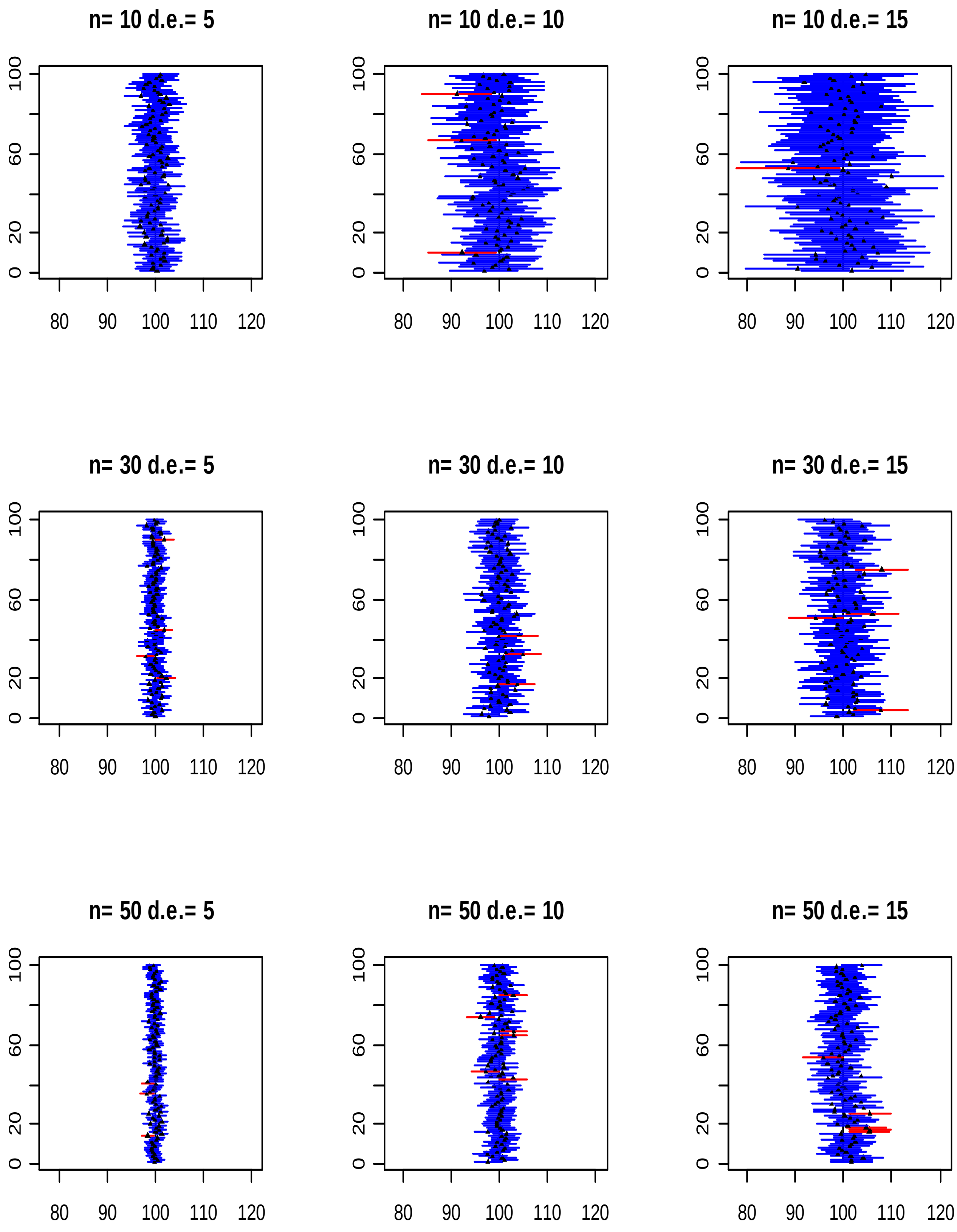

Figura 4.2. Simulación de IC del 95% para X~N (100; σ), tamaños de muestra 10, 30 y 50 y σ=5, 10 y 15. (σ=d.e.)

ESTRATEGIA DIDÁCTICA

Esta práctica requiere que el docente se apoye en una presentación en donde se exponga en detalle cada actividad y presenten para cada uno de los pasos propuestos en esta actividad las implicaciones, las soluciones analíticas y las conclusiones que se esperan. Esta propuesta complementa las discusiones que deben ocurrir con anterioridad sobre el los conceptos que rodean la distribución normal, intervalos de confianza para el promedio, el manejo de la distribución t-student y demás nociones requeridas para adelantar esta actividad.

PRESENTACIÓN ESCRITA

Para apoyar el trabajo de los estudiantes se proveerá una guía de trabajo en donde se consignen los resultados de cada grupo. Esta guía debe incorporar la plantilla en donde cada grupo entregue sus respuestas

EVALUACIÓN

Una de las características cuando se trabaja con simulación para obtener valores de variables aleatorias, es la diferencia de resultados entre los estudiantes que presentan la evaluación. Entonces las actividades contemplan dos instantes: discusión global de los procedimientos comunes involucrados y examen de las situaciones que pueden dar origen a variaciones en la respuesta. Por otra parte, es usual que cuando se desea asignar nota a una actividad como la que se propone aquí, se requiere valorar dos aspectos: el primero de ellos, es el examen de los procedimientos y cálculos que se necesitan y por otra parte las interpretaciones a lugar. En las situaciones en las que se requieren cálculos previos a las interpretaciones, se propone examinar previamente los resultados obtenidos, para lo cual es necesario revisar junto con los estudiantes que los IC que se calculan sean los correctos, buscando que las interpretaciones y conclusiones no se vean afectadas por resultados desafortunados.

Según lo anterior, las diferencias entre los estudiantes se verán reflejadas en la calidad de los argumentos y explicaciones solicitadas, en esa medida se asignará la nota respectiva al trabajo de los estudiantes.

5. Conclusiones y recomendaciones

5.1 Conclusiones

Con respecto a la revisión sobre los tópicos disciplinares se encontró que un concepto como el de IC tal y como se conoce actualmente, pasó por una serie de etapas. Cuando se enseña esta temática a los estudiantes no se comenta sobre su evolución a lo largo del tiempo y erróneamente se considera que la teoría surgió tal y como se anuncia. Se ignoran de plano las tensiones, e incluso rivalidades que se dieron entre los postulantes y sus tesis.

Los estudios que se han realizado sobre las dificultades en la comprensión de IC entre los investigadores y los estudiantes han permitido ser aun mas consientes de la complejidad de este concepto. Estudiantes e investigadores usan IC pero pasan por alto las incidencias del tamaño de muestra o la variabilidad de la característica que se está estudiando. Se encontró un fenómeno interesante: se enseña en los libros de texto así como en la práctica docente, que en la eventualidad de poder repetir la muestra en varias ocasiones, la proporción de intervalos que contiene el verdadero valor del parámetro está asociado con la confiabilidad del intervalo, pero no se advierte que esto no implica que entonces las medias de muestras sucesivas deban guardar similitud. Si se tiene la posibilidad de replicar varias veces la muestra, y con la primera se calcula un IC de 95%, no se debe esperar que en esta misma proporción las demás medias de las muestras estén contenidas dentro del primer intervalo hallado.

Ahora puedo concebir que el estudio de las dificultades sea en sí mismo un campo de investigación. En la labor docente se evidencian problemas de comprensión pero no se estudia la situación a profundidad y se argumenta de manera facilista que esto se debe enteramente a limitaciones de los educandos. Además de brindar la oportunidad de cuestionar y cuestionarse sobre las acciones propias de la docencia y del proceso de aprendizaje, conocer las dificultades en la comprensión del tema objeto de investigación, permitió apoyar las actividades de esta propuesta, incorporando situaciones y cuestionamientos para evitarlas o por lo menos hacer consientes a los estudiantes de ellas.

La propuesta está encaminada a la enseñanza de IC para la media cuando se hace simulación de muestras de una distribución Binomial y una distribución Normal, pero el enfoque puede ser usado para otros parámetros, siempre reconociendo que solo una vez se estudien a profundidad las dificultades pueda incorporarse este saber en las acciones a desarrollar en la práctica del aula.

5.2 Recomendaciones

La propuesta, aunque solo considera el estudio del promedio, puede ser extendida al estudio de otros parámetros. Por otra parte, esta propuesta solo consideró tomar muestras aleatorias de una distribución Binomial y de una distribución normal pero también pueden tomarse muestras de otras distribuciones.

Esta propuesta incorporó el estudio sobre las dificultades en relación al proceso de enseñanza-aprendizaje de loa intervalos de confianza. Esto puede ser incluso objeto de estudio, las dificultades que más fueron detalladas eran para la estimación del promedio; sobre los problemas de comprensión relativos a otros parámetros es escasa la documentación disponible.

En vista que el código en R es abierto, se puede solicitar que los estudiantes que modifiquen las rutinas ya existentes y diseñen otras aplicaciones. Por ejemplo: usar otros tamaños de muestra, cambiar las desviaciones estándar y la confiabilidad, o incluso tomar muestras aleatorias de otras distribuciones.

A. Anexo: códigos R usados

```
# Intervalo de Confianza para la media de una Distribución Binomial
# parámetros: numero de ensayos o replicas (r) y probabilidad de éxito (pe)

Simbin<-list()  #### Un archivo lista donde se almacenará para cada muestra, el promedio, límite inferior y
superior del intervalo
miu<-c()       #### Promedio de la distribución original
sigma2<-c()   #### Varianza de la variable original
sigma<-c()    #### Desviación estándar de la variable original
media<-list()
a<-0.05 ## Nivel de significancia
qnorm(1-a/2)  ### cuantil asociado a la confianza estipulada
Pe<-c(0.1,0.5,0.9,0.1,0.5,0.9,0.1,0.5,0.9) ## Probabilidades de éxito
r<-rep(8,9) ### Número de réplicas
ni<-rep(100,9) ### Numero de intervalos
n<-c(10,10,10,50,50,50,90,90,90)  ## Tamaño de la muestra

for(y in 1:length(ni)){
   miu[y]<-r[y]*Pe[y]     ### calcula miu para cada distribución
   sigma2[y]<-r[y]*Pe[y]*(1-Pe[y])    ### calcula sigma2 para cada distribución
   sigma[y]<-sqrt(sigma2[y])       ### calcula sigma para cada distribución
   Simbin[[y]]<-matrix(0,nrow=ni[y],ncol=3)
   for (d in 1:nrow(Simbin[[y]])){
     Simbin[[y]][d,1]<-mean(rbinom(n[y],r[y],Pe[y]))    ### El promedio para el intervalo
     Simbin[[y]][d,2]<-Simbin[[y]][d,1]+(qnorm(1-a/2)*(sigma[y]/(sqrt(n[y]))))    #### Limite superior del
intervalo
     Simbin[[y]][d,3]<-Simbin[[y]][d,1]-(qnorm(1-a/2)*(sigma[y]/(n[y]^0.5))) #### Limite inferior del intervalo
   }
}
#####  Creación de las graficas
par(mfrow=c(3,3))
for(i in 1:length(Simbin)){
  dta<- Simbin[[i]]
  plot(seq(min(c(dta[,3])),max(c(dta[,2])),l=100), seq(1:100), type="n", xlab=c(""), ylab=c(""))
segments(miu[i],1,miu[i],100)
for (f in 1:nrow(dta)){
  if (miu[i]>dta[f,3] & miu[i]<dta[f,2]){
    segments(dta[f,3],f,dta[f,2],f, col="blue")
  }else (
```

```
    segments(dta[f,3],f,dta[f,2],f, col="red")
    )
}
}



## Intervalo de confianza para muestras de una distribución normal
## Promedio = 100 y variando el tamaño de muestra y la desviación estándar.

ni<-rep(100,9)      #numero de nuestras a elegir
n<-c(10,10,10,30,30,30,50,50,50)      #tamaño de cada una de las muestras
sigma<-c(5,10,15,5,10,15,5,10,15)
MED<-100

Simnorm<-list()
for (o in 1:length(n)){
  Simnorm[[o]]<-matrix(0,nrow=ni[o],ncol=3)
  colnames(Simnorm[[o]])<-rep(paste("n=",n[o],"Sig=",sigma[o]),3)
  for (d in 1:nrow(Simnorm[[o]])){
  Simnorm[[o]][d,1]<-mean(rnorm(n[o],MED,sigma[o]))
  Simnorm[[o]][d,2]<-Simnorm[[o]][d,1]-((qt(0.975,n[o]-1))*(sigma[o]/(sqrt(n[o]))))
  Simnorm[[o]][d,3]<-Simnorm[[o]][d,1]+((qt(0.975,n[o]-1))*(sigma[o]/(n[o]^0.5)))
  }
}


par(mfrow=c(3,3))
for(i in 1:length(Simnorm)){3
  dta<- Simnorm[[i]]
  plot(seq(min(unlist(Simnorm)),max(unlist(Simnorm)),l=100), seq(1:100), type="n", xlab=c(""), ylab=c(""),
main=names(Simnorm[[i]][d,3]))
segments(MED,1,MED,100)
for (f in 1:nrow(dta)){
  if (MED>dta[f,2] & MED<dta[f,3]){
    segments(dta[f,3],f,dta[f,2],f, col="blue")
  }else (
    segments(dta[f,3],f,dta[f,2],f, col="red")
    )
points(dta[f,1], f, pch=17, col="black", cex=0.5)
}
}
```

Bibliografía

Behar, R., & Grima, P. (2001). Mil y una dimensiones del aprendizaje de la estadística. *Estadística Española , 43* (148), 189-207.

Campanario, J. M., & Moya, A. (1999). ¿Como enseñar ciencias? Principales tendencias y propuestas. *Enseñanza de las ciencias , 17* (2), 179-192.

Canavos, G. (1988). *Probabilidad y Estadística. Aplicaciones y Métodos.* México: Mcraw-Hill.

Cumming, G. (2012). *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.* New York: Routledge.

Cumming, G., & Fidler, F. (2005). Interval estimates for statistical communications: problems and possible solutions. *IASE Satellite Conference on Statistics Education and the Communication of Statistics.* Sydney: International Association for Statistical Education.

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

delMas, R., Garfield, J., & Chance, B. (1999). Exploring the role of computer simulations in developing understanding of sampling distributions. *Annual Meetings of the American Educational Research Association.* Montreal, Quebec.

delMas, R., Garfield, J., & Chance, B. (2004). Using assessment to study the development of students´ reasoning about sampling distribution. *Annual Meetings of the American Educational Research Association.* California.

Devore, J. (2008). *Probabilidad y Estadística para Ingenería y Ciencias.* México: Cengale Learning.

Díaz, C. (2007). *Viabilidad de la enseñanza de la inferencia bayesiana en el analisis de datos en psicologia.* Tesis Doctoral, Universidad de Granada.

Estes, W. K. (1997). On the communications of information by displays of standard errors and confidence intervals. *Psychonomic Bulletin & Review , 4* (3), 330-341.

Euclides. (1991). *Elementos.* (M. Puertas, Trad.) Madrid: Gredos.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research. *Journal for Research in Mathematics Educations , 19* (1), 44-63.

Gutiérrez Cabria, S. (1994). *Filosofía de la Estadística.* Valencia: Universitat de Valencia.

Kempthorne, O. (1980). The teaching of statistics: content versus form. *The American Statistician , 34* (1), 17-21.

Mayo, D. (1981). In defense of the Neyman-Pearson theory of confidence intervals. *Philosophy of Science , 48* (2), 269-280.

Olivo, E., Batanero, C., & Díaz, C. (2007). Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Edicación Matemática , 20* (3), 5-32.

Olivo, E., Ortiz, J. J., & Batanero, C. (2007). Notas historicas sobre los intervalos de confianza e implicaciones didácticas. *X Simposio de la SEIEM* (págs. 105-114). Huesca: Sociedad Española de Investigación en Educación Matemática.

Rao, R. (1997). *Statistics and Truth: Putting Chance to Work.* Singapore: World Scientific Publishing.

Salas, C. (2008). ¿Por qué comprar un paquete estadístico si existe R? *Ecología Austral , 18* (2), 223-231.

Wackerly, D., Mendenhall, W., & Scheaffer, R. (2002). *Estadística Matemática con Aplicaciones.* Mexico: Thomson.

Walpole, R., Myers, R., Myers, S., & Ye, K. (2007). *Probabilidad y Estadística para Ingeniería y Ciencias.* Mexico: Pearson Educación.

Yañez, S. (2000). La estadística una ciencia del siglo XX. R.A. Fisher, el genio. *Revista Colombiana de Estadística , 23* (2), 1-14.