INTERVALOS DE CONFIANZA, DESARROLLO HISTÓRICO E IMPLICACIONES DIDÁCTICAS

Eusebio Olivo[1], Carmen Batanero[2]
[1]Instituto Tecnológico de Estudios Superiores de Monterrey. México
[2]Departamento de Didáctica de las Matemáticas, Universidad de Granada
eusebio.olivo@itesm.mx, batanero@ugr.es
Reporte de Investigación

Resumen

En este trabajo analizamos el desarrollo histórico de la evolución de las ideas relacionadas con la estimación, que culminan con el concepto intervalos de confianza, con la finalidad de identificar algunos de sus campos de problemas, así como dificultades de tipo epistemológico que puedan repetirse en el aprendizaje de los alumnos. Finalizamos con unas conclusiones para mejorar la enseñanza del tema.

Palabras Clave: Intervalos de confianza; Origen histórico; Probabilidades Inversas; Probabilidad Fiducial; Intervalos de Credibilidad; Implicaciones didácticas

Introducción

El intervalo de confianza es un tema estudiado en todos los cursos de estadística universitarios e incluso en la educación secundaria. Por ejemplo, en España, en M.E.C. (2004), se introduce dentro del tema de *estimación* en el Bachillerato de Ciencias Sociales.

Por otro lado, las investigaciones psicológicas y las didácticas han avisado acerca de errores en la inferencia estadística, sobre todo en la interpretación del contraste de hipótesis (Morrison y Henkel, 1970; Vallecillos, 1994, 1999; Harlow, Mulaik y Steiger, 1997; Batanero, 2000; Díaz y de la Fuente, 2004). Al mismo tiempo, investigadores de prestigio en la comunidad científica, sugieren el uso de los intervalos de confianza para complementar los contrastes de hipótesis y mejorar de este modo los errores denunciados en la práctica de la inferencia estadística (e.g. Davies, 1998; Clark, 2004).

Este cambio metodológico, requiere asegurar que las dificultades sobre los tests de hipótesis no se repiten –o al menos no con tanta intensidad- en los intervalos de confianza, tema donde la investigación didáctica es todavía incipiente. Algunos trabajos vinculados son los de Cumming, William y Fidler (2004), relacionado con errores que cometen los investigadores y Behar (2001) y Terán (2006), quienes inician el estudio de los errores de los estudiantes, relacionados sobre todo con la interpretación del coeficiente de confianza y de los extremos del intervalo.

La indagación del origen histórico del intervalo de confianza, nos permitirá identificar algunos de sus campos de problemas, que es uno de los elementos considerados en el marco teórico en el cual nos basamos; el modelo teórico sobre el significado de los objetos matemáticos (Godino y Batanero, 1998; Godino, 2002). Este trabajo está inscrito en un estudio más amplio que tiene como objetivo general analizar el significado presentado sobre los Intervalos de Confianza en los libros de texto utilizados en la formación de ingenieros en México. Más específicamente, y utilizando el enfoque ontosemiótico de la cognición matemática (EOS) propuesto por Godino y colaboradores (Godino y Batanero, 1994; Godino, 2002; Godino, 2003), en ese estudio presentamos un análisis y clasificación de los *elementos de significado* de la noción de intervalo de confianza en dicho contexto. Dicho estudio se apoya en el análisis del significado actual del concepto y su evolución histórica, así como en el marco teórico citado y revisión bibliográfica de investigaciones previas. Los resultados obtenidos serán la base de la posterior construcción de un cuestionario de evaluación de los significados personales que asignan los estudiantes a este objeto matemático y servirán también para fundamentar futuras propuestas de enseñanza.

El enfoque teórico denominado "enfoque ontosemiótico" (EOS) de la cognición matemática proporciona una perspectiva pragmática – antropológica sobre el conocimiento matemático que puede complementar y enriquecer el análisis que hacemos de la evolución histórica del término. El EOS propone tres dimensiones en el análisis de la enseñanza y el aprendizaje de las matemáticas: epistemológica, cognitiva e instruccional. Cada una de ellas se aborda con herramientas agrupadas en tres modelos teóricos: teoría de los significados institucionales y personales de los objetos matemáticos (Godino y Batanero, 1994, 1998); teoría de las funciones semióticas (Godino, 2002; Godino, Batanero y Roa, 2005) y teoría de las configuraciones didácticas (Godino, Contreras y Font, 2006).En ese estudio nos ocupamos principalmente de la dimensión epistemológica.

El punto de partida del EOS es una ontología de objetos matemáticos que tiene en cuenta el triple aspecto de la actividad matemática de resolución de problemas: socialmente compartida, lenguaje simbólico y sistema conceptual lógicamente organizado. Tomando como noción primitiva la de situación-problemática, se definen los conceptos teóricos de práctica, objeto (personal e institucional) y significado, con el fin de hacer patente y operativo, este triple carácter de la matemática así como la génesis personal e institucional del conocimiento matemático.

En lo que sigue presentamos el análisis del significado actual del intervalo de confianza, y un estudio histórico de la evolución del concepto, como primer paso para continuar las investigaciones anteriores y analizar los posibles errores de aprendizaje de los estudiantes. El análisis conceptual que realizamos de los intervalos de confianza es desde un punto de vista estructural, esto es, teniendo en cuenta los elementos conceptuales y procedimentales puestos en juego en nuestro objeto de estudio, que el estudiante ha de comprender de antemano para abordar su estudio, lo que le dota de complejidad semiótica.

Complejidad del significado del intervalo de confianza

El intervalo de confianza se puede considerar como un concepto y como un procedimiento. Considerado como concepto, y en el contexto de estimar un parámetro poblacional, un intervalo de confianza es un rango de valores (calculado a partir de los datos de una muestra) en el cual podría encontrarse el verdadero valor del parámetro, junto con un coeficiente de confianza que indica el porcentaje de muestras tomadas en las mismas condiciones, en las cuales el intervalo cubriría el verdadero valor del parámetro. Como procedimiento, da una regla general de construcción de dicho rango de valores a partir de un estadístico calculado en los datos de la muestra, para el parámetro correspondiente. La idea general de intervalo de confianza se particulariza dependiendo del parámetro a estimar (media, proporción, varianza, etc.) y según las condiciones (tipo de distribución, qué se conoce de la misma, etc.).

La descripción resumida e intuitiva anterior se desarrolla en el mapa conceptual representado en la figura 1, donde observamos la complejidad del concepto, que se apoya en muchos otros y donde los mismos conceptos asociados aparecen a diferentes niveles. Por ejemplo, la idea de distribución aparece relacionada con la variable aleatoria (distribución de la población), con la muestra (distribución estadística de datos) y con el muestreo (distribución muestral del estadístico).

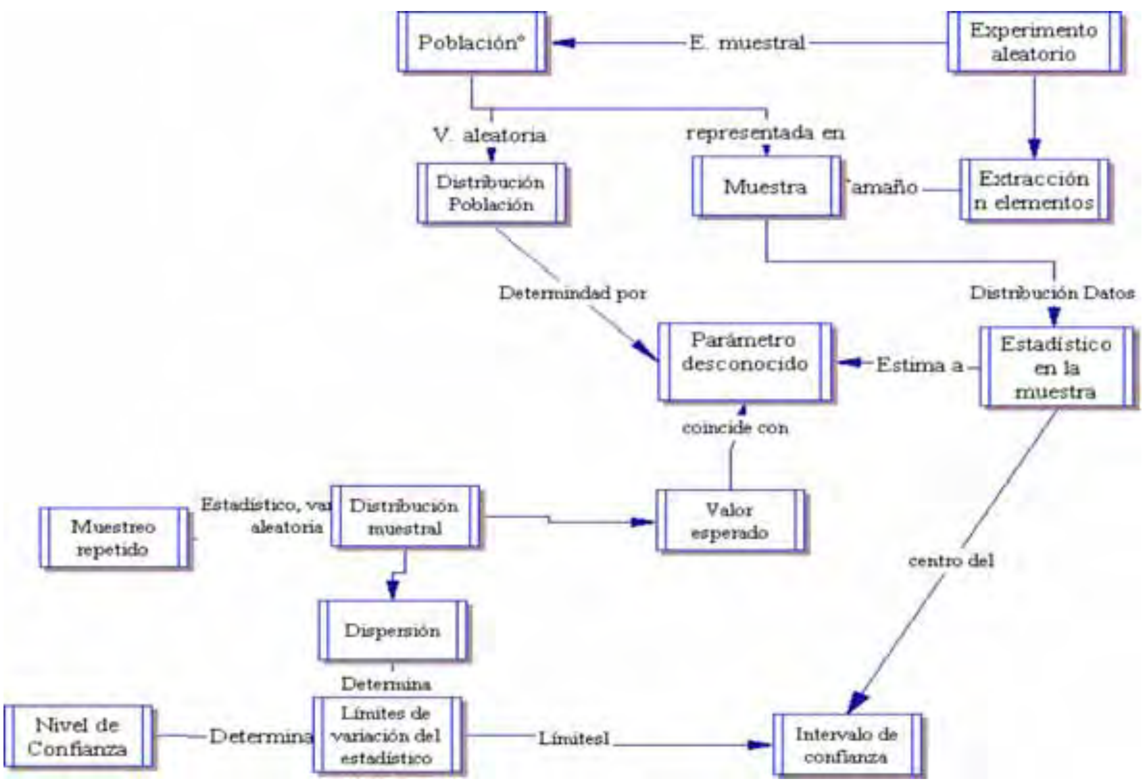

Figura 1. Mapa conceptual del Intervalo de Confianza

Origen histórico

La estimación estadística, hasta el siglo XX, se basaba fundamentalmente en el método de mínimos cuadrados debido a Gauss y el método de desviación mínima absoluta ideado por Laplace, que eran usados generalmente para estimar parámetros en modelos lineales (Rao, 1992, p.36). Desde el punto de vista de la filosofía de la ciencia, el problema de la estimación se relaciona con la inferencia inductiva, es decir aquella forma de razonamiento según la cual la verdad de las premisas no comporta necesariamente la de la conclusión, y en términos estadísticos, como las argumentaciones de la muestra hacia la población de la cual, se extrajo dicha muestra (Vallecillos, 1994).

Matemáticamente, el problema de la estimación, podemos plantearlo de distintas maneras. Una de ellas es asumir un fenómeno aleatorio que viene caracterizado por una distribución de probabilidad, que depende de uno o varios parámetros, supuestos *constantes*. Al no ser posible recolectar los datos de toda la población, hemos de conformarnos con una muestra aleatoria, de la misma población. El problema formulado es dar un valor aproximado del parámetro (o parámetros) a partir de los datos observados del estadístico (o estadísticos) en la muestra.

Es decir, siendo $X_1, X_2, ..., X_n$ un sistema de n variables aleatorias, cuyos valores particulares pueden ser obtenidos a través de muestras aleatorias. Como la ley de probabilidad de estas variables

$$P(x_1, x_2, ..., x_n \mid \theta_1, \theta_2, ..., \theta_k)$$

para $x_1, x_2, ..., x_n$ valores de las n variables respectivamente, dependen de k parámetros desconocidos, se trata de estimar estos parámetros haciendo uso en este proceso de los valores observados $x'_1, x'_2, ..., x'_n$ del sistema de n variables aleatorias.

Probabilidadades inversas

Thomas Bayes (1763), en un ensayo póstumo, propone una respuesta al problema de encontrar la probabilidad de que un efecto ocurrido sea debido a una causa dada, tomando como punto de partida las probabilidades a priori, de las posibles causas,
De este modo, plantea el uso de la probabilidad matemática para justificar la inferencia inductiva. Como veremos, el Teorema de Bayes constituye el primer esfuerzo de solución del problema de la estimación por intervalos, aunque considerando que los parámetros $\theta_1, \theta_2, ..., \theta_k$ son *variables aleatorias,* caracterizados por una distribución a priori de probabilidad que se actualiza mediante el producto por las verosimilitudes. (Rivadulla 1991). El desarrollo es como sigue

Si en una serie de experimentos un suceso ha aparecido p veces y q veces el suceso contrario, la posibilidad w de que la probabilidad buscada para el suceso se encuentre entre x y X viene dada por la siguiente expresión:

$$W = \frac{\int_x^X x^P (1-x)^q \, dx}{\int_0^1 x^P (1-x)^q \, dx}$$

Puesto que Bayes considera el parámetro como variable aleatoria (al igual que hace hoy día la escuela Bayesiana), al dar la probabilidad de que el parámetro se encuentre entre ciertos valores, en realidad está dando un avance hacia la construcción de *intervalos de credibilidad*. Estos intervalos se utilizan en inferencia bayesiana y en ellos se consideran los extremos como fijos y el parámetro como aleatorio. Los *intervalos bayesianos de credibilidad* serían entonces anteriores históricamente a los *intervalos de confianza.*

Laplace rescata el teorema de Bayes y desarrolla él mismo algunas de sus consecuencias. En "*Mémoire sur la Probabilité des Causes par les événements"* publicado por Laplance en 1774 encontramos también un análisis Bayesiano, donde se ocupa de la inferencia, con la finalidad de determinar la proporción desconocida (parámetro de una distribución binomial) (Hald, 1998, pp. 23-4). Asimismo, en la sección V del "Mémoire sur la Probabilité des Causes par les événements", Laplace se enfoca en la estimación del valor medio a partir de tres observaciones. Motivado por una nota publicada por J. Bernoulli III que indicaba que el problema de estimación de la media era de considerable interés para los astrónomos. La importancia que guarda el artículo citado de Laplace radica en que sienta las bases de la teoría de la decisión moderna y de la inferencia bayesiana (Stigler 1986). En 1818 Laplace establece la primera formulación del problema de *estimación puntual.*

Tanto Laplace como, posteriormente Gauss (hacia 1887) contemplan un valor desconocido del parámetro θ y un cierto número de sus mediciones x_i todas sujetas a un error aleatorio. También los dos contemplan la formulación de lo que se llama la "función pérdida" $L(\hat{\theta},\theta)$.Esta función representa el error que el estadístico asumirá por adoptar a $\hat{\theta}$ como estimador de θ. Laplace usó el valor absoluto de la diferencia $L_L(\hat{\theta},\theta) = \left|\hat{\theta}-\theta\right|$, mientras que Gauss prefirió su cuadrado $L_G(\hat{\theta},\theta) = (\hat{\theta}-\theta)^2$. De ahí resultó la teoría de mínimos cuadrados. (Neyman, 1976).

Probabilidad fiducial

Fisher (1930) introduce la idea de *intervalos fiduciarios,* para tratar de salvar las dificultades que surgen al intentar usar el Teorema de Bayes cuando no existe información a priori sobre los parámetros. Fisher, que siempre se preocupó por mantener una visión frecuencial de la probabilidad desarrolló el método llamado "fiducial" basado en la función de verosimilitud (Rouanet, 1998). La probabilidad fiducial trata de expresar la frecuencia con que el valor verdadero de un parámetro toma un valor determinado; a partir de los datos observados, por ejemplo que la probabilidad de que cierto parámetro poblacional sea menor que un valor dado es del 5%.

Por ejemplo, para una población normal con σ conocida, tomemos el valor dado $\mu_1 = \overline{X} - 1.64\frac{\sigma}{\sqrt{n}}$. Lo que el enunciado de probabilidad fiducial afirma es que $P(\mu < \mu_1) = 0.05$, lo cual se interpreta como que en el 5 % de todas las posibles muestras, μ no alcanzarán el valor μ_1.

En el caso general, si $\hat{\theta}$ es un estadístico y P la probabilidad de que $\hat{\theta}$ sea menor que un valor específico, entonces tenemos una relación de la forma

$$P = F(\hat{\theta}, \theta)$$

Si ahora damos a P cualquier valor particular tal como 0.95, tenemos una relación entre el estadístico $\hat{\theta}$ y el parámetro θ, tal que $\hat{\theta}$ es el valor en el percentil 95 % correspondiente a un θ dado, esta relación implica que en el 5% de las muestras $\hat{\theta}$ excederá al percentil 95% correspondiente al valor real de θ de la población de la cual fue extraída. Fisher (1930) llama a esa relación el *"5% fiducial del valor de θ"* correspondiente a un valor dado de $\hat{\theta}$. Si $\hat{\theta}$ se incrementa con θ para todos los valores posibles, debemos expresar la relación diciendo que "*el verdadero valor de θ será menor que el 5% fiducial del valor correspondiente al valor observado de $\hat{\theta}$ en exactamente 5 intentos de 100*" (Fisher, 1930).

Este método trata de evitar el uso de probabilidades a priori para el parámetro (como la estadística frecuencial), pero produce probabilidades a posteriori del parámetro, dados los datos (como la inferencia bayesiana). Rouanet (1998) indica que en algunos casos las distribuciones fiduciales de Fisher coinciden con las distribuciones bayesianas a posteriori, de modo que se podría considerar que Fisher fue un bayesiano, sin saberlo. Sería el caso de distribución inicial no informativa, es decir, cuando se suponen equiprobables a priori todos los valores de los parámetros.

Para Fisher uno de los objetivos de la investigación empírica debe ser la búsqueda de la máxima verosimilitud. Esta idea da lugar a los métodos de estimación de máxima verosimilitud. (Rivadulla, 1991). Fisher con sus constructos de verosimilitud y probabilidad fiducial trata de sustituir la probabilidad subjetiva, aportada por el método bayesiano, por una medida de creencia racional, basada únicamente en los datos observados, pero continúa argumentando sobre una población tomando como punto de partida los datos recolectados en el muestreo.

Las teorías de la probabilidad fiducial de Fisher permiten calcular valores de verosimilitud sobre los parámetros, pero no proporcionan una distribución de probabilidad acerca de parámetros desconocidos, por lo que no resultaron exitosos. Aún así, hoy día la idea de verosimilitud y máxima verosimilitud es una de las principales en estimación. Por otro lado los métodos bayesianos no informativos se apoyan en el método fiducial de Fisher (Lecoutre, 1999).

Experimentos repetidos e intervalos de confianza

Jerzy Neyman (1894-1981) inicia la teoría moderna de *intervalos de confianza* e introduce el término en 1934 con su artículo, "On the two different aspects of the representative method", arrancando con ello la teoría estadística de la estimación del programa Neyman- Pearson. Pero su trabajo más importante sobre el problema de la estimación estadística es "Outline of a theory of statistical estimation based on the classical theory of probability " de 1937.
Neyman (1934) muestra que el problema de la estimación puede ser resuelto con base en la teoría frecuentista de probabilidades y sin requerir algún conocimiento de probabilidades a priori. Para ello sugiere que la solución del problema de estimación consiste en determinar ciertos intervalos, que denomina intervalos de confianza, en los cuales debemos asumir están contenidos los valores de los parámetros; la probabilidad de un error será igual o menor que $1-\varepsilon$ donde ε es cualquier número $0<\varepsilon<1$, escogido anticipadamente.

El número ε lo llama coeficiente de confianza. La forma de interpretar esta probabilidad es la siguiente: al estimar el valor del parámetro θ, el experimentador o estadístico tendrá éxito, en *aproximadamente un porcentaje de muestras tomadas de la misma población* a la larga. Es decir, en el intervalo de confianza el parámetro se considera constante, pero los extremos son variables aleatorias que cambian de muestra a muestra. Su método, busca la máxima exactitud posible del resultado al determinar las estimaciones inferior y superior, obtenidas al sumar y restar al estimador la desviación respecto al estimador.

Neyman (1941) intenta probar que no hay relación entre la teoría fiducial y la teoría de los intervalos de confianza, aunque pasa por serias dudas provocadas, entre otras cosas, por la identidad numérica de los límites fiduciales de Fisher con los límites de sus intervalos de confianza. También, para el caso de distribución a priori no informativa los intervalos de credibilidad y confianza coinciden (sus límites) en muchos casos, aunque la interpretación, como hemos visto es muy diferente (Lecoutre, 1999).

Algunas conclusiones

El problema de estimación por intervalos se resolvió históricamente por tres métodos diferentes: los intervalos de credibilidad (solución bayesiana), los intervalos fiduciales de Fisher y los intervalos de confianza de Neyman.

Las teorías de la probabilidad fiducial de Fisher permiten calcular valores de verosimilitud sobre los parámetros, pero no proporcionan una distribución de probabilidad acerca de parámetros desconocidos, por lo que no resultaron exitosos. Aún así, hoy día la idea de verosimilitud y máxima verosimilitud es una de las principales en estimación. Por otro lado los métodos bayesianos no informativos se apoyan en el método fiducial de Fisher (Lecoutre, 1999).

Respecto a la diferencia entre intervalos de credibilidad y de confianza, incluso aunque en algunos casos coincida numéricamente el valor de sus extremos, la historia nos revela que tienen una interpretación muy diferentes:

- En el intervalo de confianza el parámetro es constante y los extremos del intervalo son aleatorios. El nivel de confianza $(1-\alpha)100\%$ significa para la teoría Neyman-Pearson que a la larga el $(1-\alpha)100\%$ de los intervalos (calculados en muestras sucesivas de la misma población) incluyen el valor verdadero del parámetro θ que se desea estimar. La probabilidad $(1-\alpha)100\%$ es una probabilidad frecuencial y se refiere al experimento supuesto de repetir indefinidamente la toma de muestras de la misma población y calcular los intervalos.

- La teoría Bayesiana considera θ como una variable aleatoria, con una distribución inicial de probabilidad. El teorema de Bayes actualiza esta distribución inicial pasando a una distribución final de probabilidades para el parámetro. No se plantea el contexto de repetición. Los límites de un intervalo de credibilidad de $(1-\alpha)100\%$ se consideran fijos y la probabilidad es una probabilidad subjetiva y epistémica, porque se refiere sólo al experimento concreto.

De hecho (Cumming, William y Fidler, 2004; Belia, Fidler y Cumming, 2005; Schenker y Gentleman, 2001), muchos investigadores dan una interpretación bayesiana a los intervalos de confianza, es decir, piensan que el intervalo de confianza da la probabilidad de que el verdadero valor del parámetro θ está contenido dentro de los límites de confianza (esta probabilidad la daría el intervalo de credibilidad).

Puesto que el estudio histórico muestra que el desarrollo de las ideas bayesianas sobre estimación fue bastante anterior al de las ideas frecuenciales (aunque no se llegaron a imponer en la práctica por la dificultad filosófica de tener que trabajar con probabilidades subjetivas), una hipótesis es que el método Bayesiano podría ser más intuitivo que el frecuencial. Esto es también sugerido por algunos investigadores (Lecoutre, 1999) aunque habría que analizarlo con más detalle en nuevas investigaciones.

Reconocimientos

Agradecimiento a la Fundación Carolina (Madrid), Tec de Monterrey, Campus Monterrey y Proyecto SEJ2004-00789, MEC (Madrid).

Referencias Bibliográficas

Batanero, C. (2000). Controversies around significance tests. *Journal of Mathematics Thinking and Learning,* 2(1-2), 75-98.

Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística.* Ph. D. Universidad Politécnica de Cataluña.

Clark, M. L. (2004). Los valores p y los intervalos de confianza, ¿en qué confiar?. *Revista Panamericana de Salud Pública,* 15(5), 295-296.

Cumming, G. y Fidler, F. (2005). Interval estimates for statistical communication: problems and possible solutions. Trabajo presentado en la *IASE/ISI Satellite Conference on Communication of Statistics.* Sydney: IASE.

Cumming, G. y Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *American Psychologist, 60,170-180*

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

Davies, H. (1998). What are confidence intervals? On line: [http://www.evidence-based-medicine.co.uk]

Díaz, C. y de la Fuente, I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento*. Volumen especial 2004, 161-167.

Fisher, R. A. (1930). Inverse Probability. *Proceedings of the Cambridge Philosofical Society,* 26, 528-535.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques,* 22 (2/3), 237-284.

Godino, J. D. (2003). Teoría de las funciones semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática. *Departamento de Didáctica de la Matemática. Universidad de Granada*. Disponible en Internet: URL: http://www.ugr.es/local/jgodino/indice_tfs.htm.

Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathematiques*, 14(3), 325-355.

Godino, J. D. y Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education. En A. Sierpinska y J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.

Godino, J. D., Batanero, C. y Roa R. (2005). Análisis onto-semiótico de problemas combinatorios y de su resolución por estudiantes universitarios. *Educational Studies in Mathematics,* 60(1), 3-36

Harlow, L. L.; Mulaik, S. A. y Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

Hald, A. (1998). *A history of mathematical statistics from 1750 to 1930*. New York: John Wiley.

Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tome 58, Book 2) (pp. 205 – 208). Helsinki, Finland: International Statistical Institute.

Morrison, D.E. y Henkel, R.E. (1970). *The significance test controversy*. Chicago: Aldine.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. Royal Society of London*, series A, 767.

Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biométrica* 32, 2, 128-150.

Neyman, J. (1976). The emergence of mathematical statistics En D. B. Owen (Ed.), *On the history of statistics and probability* (pp. 149-189). New York: Marcel Dekker, Inc.

Rao, C. R. (1992). R. A. Fisher: The founder of modern statistics. *Statistical Science*, 7, 1, 34-48

Rivadulla, A. (1991). *Probabilidad e Inferencia científica.* Barcelona: Anthropos.

Rouanet, H. (1998a). Statistics for researchers. En H. Rouanet et al. (Eds.), *New ways in statistical methodology* (pp. 1 – 28). Berna: Peter Lang.

Stigler, S. M. (1986). *The history of statistics the measurement of uncertainty before 1900*. Prensa de Belknap: Universidad de Harvard.

Terán, T. (2005). Elements of meaning and its role in the interaction with a computacional program. Trabajo contribuido en A. Rossman y B. Chance (Eds.), *Proceedings of ICOTS 7.* Salvador (Bahia), Brasil: IASE. CD ROM.

Vallecillos, A. (1994). *Estudio teórico - experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios.* Tesis doctoral Departamento de Didáctica de la Matemática. Universidad de Granada.

Vallecillos, A. (1999). Some empirical evidences on learning difficulties about testing hyphoteses. Ponencia invitada. *Proceedings of the 52nd Session of the ISI* (Vol.2, 2001-204). The Netherland:ISI.