



Comparación de dos métodos para el cálculo de la incertidumbre en los análisis de laboratorio

J. Llorca¹ / T. Dierssen Sotos²

¹Cátedra de Medicina Preventiva y Salud Pública. Facultad de Medicina. Universidad de Cantabria.

²Servicio de Medicina Preventiva. Hospital de Sierrallana (Torrelavega).

Correspondencia: Dr. Javier Llorca Díaz. Cátedra de Medicina Preventiva y Salud Pública. Facultad de Medicina de la Universidad de Cantabria. 39011 Santander. e-mail: llorcaj@medi.unican.es

Recibido: 19 de octubre de 1999

Aceptado: 22 de agosto de 2000

(Comparing two methods in the estimation of the uncertainty in laboratory quality)

Resumen

Objetivo: Comparación de dos métodos para el cálculo de incertidumbres en el control de calidad de los laboratorios.

Métodos: Mediante una simulación por ordenador, se comparan el método delta (propuesto por la International Organization for Standardization y por la Entidad Nacional de Acreditación) y un método basado en el *bootstrap*. La simulación incluye varias situaciones en las que las condiciones ambientales y la relación entre las variables medidas se van modificando.

Resultados: El porcentaje medio de cobertura de los intervalos de confianza al 95% es ligeramente más alto y más próximo al nominal en el *bootstrap* que en el delta. Las mayores diferencias se aprecian en la distribución de los porcentajes de cobertura: mientras en el *bootstrap* un gran número de simulaciones tienen porcentajes de cobertura próximos al 95%, en el delta los porcentajes obtenidos ofrecen mucha mayor dispersión, llegando en ocasiones al 100% de cobertura o descendiendo hasta el 80% o incluso menores. El *bootstrap* ofrece resultados superponibles, incluso cuando intervienen variables desconocidas que no han podido ser medidas o cuando las variables consideradas están correlacionadas. En cambio, el método delta ofrece resultados más pobres en estas dos circunstancias.

Conclusión: La incertidumbre en el control de calidad de los laboratorios puede ser obtenida por *bootstrap* con ventaja sobre el método delta.

Palabras clave: Control de calidad. Incertidumbre. Error estándar. *Bootstrap*. Método delta.

Abstract

Objective: To compare two methods in the estimation of the uncertainty in laboratory quality control.

Methods: A computerized simulation is performed to compare the delta method (suggested by the International Organization for Standardization and the Entidad Nacional de Acreditación) and a bootstrap-based method. The simulation includes several situations with different environmental conditions and different relationships between the analyzed variables.

Results: The mean in the coverage obtained by the estimated confidence intervals is higher and closer to the nominal using the bootstrap than using the delta method. The most important differences are observed in the coverage percent distribution: while using the bootstrap, a great number of simulations obtain coverages near the nominal of 95%; using the delta method the coverages are more sparsed, including coverages of 100% in some occasions and lesser than 80% in others. The bootstrap offers very similar results under different conditions, including in the presence of unknown and unmeasured variables or when the analyzed variables are correlated. The delta method shows poorer results in both situations.

Conclusion: The uncertainty in the laboratory quality control can be estimated more accurately with bootstrapping than with the delta method.

Introducción

El control de calidad de los servicios sanitarios es una tarea a la que la salud pública está dedicando una atención creciente: en los últimos años se ha hecho un esfuerzo para acreditar protocolos o extender guías de práctica clínica. El movimiento de Medicina Basada en la Evidencia, por otra parte, intenta

que la toma de decisiones clínicas se fundamente en la mejor evidencia científica disponible.

Las determinaciones que realizan los laboratorios juegan un papel fundamental tanto en la industria (por sus consecuencias económicas) como en la sanidad (por las decisiones clínicas que sustentan y sus consecuencias). Para garantizar la fiabilidad de los resultados se han desarrollado sistemas de calidad basa-

dos en la norma europea EN 45001 o en los principios de Buenas Prácticas del Laboratorio de la OCDE.

La acreditación de laboratorios tiene como finalidad garantizar que los análisis sean reproducibles y se lleven a cabo mediante métodos correctos y con equipamiento adecuado. Este es un punto crucial en el control de calidad de los centros sanitarios. La Unión Europea ha establecido la obligatoriedad de que los laboratorios dedicados al análisis de alimentos con destino humano estén acreditados por los organismos de acreditación nacionales, de acuerdo con la norma de la International Organization for Standardization (ISO). En España esta acreditación se lleva a cabo por la Entidad Nacional de Acreditación (ENAC), la cual ha comenzado también a acreditar laboratorios clínicos desde 1998.

Un elemento fundamental en el proceso de acreditación es el cálculo de incertidumbres. Supongamos, por ejemplo, que determinado ensayo requiere un cultivo en estufa a 37 ± 1 °C; ¿en qué medida puede el analista garantizar que el ensayo se ha realizado en estas condiciones? La medición de la incertidumbre pretende dar respuesta a esta pregunta. La ISO, junto con otras organizaciones internacionales de medida, presentó en 1993 una guía (en adelante, la guía ISO) para el cálculo de incertidumbres, en la que se señala que la desviación típica es un estimador de la incertidumbre de la magnitud medida¹. En un documento previo (WECC-19-1990)² hay cierta confusión en la terminología entre desviación típica de la distribución de medidas y desviación típica de la distribución de las medias (también conocida como el error estándar de la media), aunque un análisis cuidadoso de las fórmulas 4, 9 y 10 del documento deja claro que se está calculando el error estándar de la media. La confusión fue corregida en la revisión que publicó EAL en 1997 con el título *Expression of the Uncertainty of Measurement in Calibration*³, traducido al español en 1998 por ENAC⁴: la incertidumbre en una calibración se calcula como un error estándar de la media.

En este artículo se revisa la manera en que la guía ISO propone realizar el cálculo de incertidumbres (conocida como método delta) y se sugiere un método alternativo empleando el *bootstrap*. Para comparar ambos métodos se realiza una simulación por ordenador que muestra el comportamiento del método delta y del *bootstrap* bajo diferentes condiciones.

Método

La incertidumbre en los documentos de la ISO

Ilustraremos tanto este apartado como los siguientes analizando el ejemplo de la incertidumbre de la temperatura en una estufa de cultivo. En este caso la tem-

peratura real de la estufa depende de la uniformidad (variación de la temperatura tomada simultáneamente en diferentes puntos de la estufa) y la estabilidad (variación de la temperatura a lo largo del tiempo en un solo punto). En cambio, la temperatura medida al leer el dial de la estufa depende de la uniformidad, la estabilidad, la división de escala del termómetro y la incertidumbre del mismo termómetro.

Los documentos de la ISO proponen emplear el siguiente método para el cálculo de incertidumbres:¹⁻⁴

1) Se identifica la magnitud a medir (en nuestro caso la temperatura) como una función de varias componentes. En nuestro caso, $T_m = T_t + U + E + P + D$, donde T_m es la temperatura medida, T_t es la temperatura prevista por el termostato, U es el error introducido por la uniformidad, E es el error introducido por la estabilidad, P es el error provocado por la variabilidad del termómetro y D es el error inducido por la división de escala del termómetro.

2) Se desarrolla la fórmula anterior en una serie de Taylor de primer orden. Una serie de Taylor es un procedimiento matemático que permite calcular de manera aproximada el valor de una función f en un punto x como combinación del valor de la función y de sus derivadas en otro punto a : $f(x) \approx f(a) + f'(a)(x-a)$ ⁵.

3) Por procedimientos algebraicos se sitúa en el lado izquierdo de la ecuación el error estándar de la T_m , que queda como una función de los errores estándar de la uniformidad, la estabilidad, el termómetro patrón y el error de la división de escala: $l = (\sigma_u^2/n_u + \sigma_e^2/n_e + \sigma_p^2/n_p + \sigma_d^2)^{1/2}$, donde σ^2 indica la varianza de cada variable y n indica el número de lecturas realizadas para su cálculo (se ha asumido $n_d = 1$).

Este procedimiento es conocido como el método delta⁶, fue desarrollado en los años cuarenta y, hasta la aparición de los ordenadores, ha sido el método más empleado para estimar varianzas cuando no había un método directo que permitiera su cálculo. Del desarrollo anterior deben notarse algunos aspectos relevantes: En primer lugar, se asume que todos los componentes que influyen en la temperatura son conocidos y mensurables. En segundo lugar, se ha asumido que la relación entre la temperatura y sus componentes es lineal, lo que permite que los términos del desarrollo de Taylor de orden mayor que uno sean cero. Por último, siguiendo las recomendaciones de la guía ISO, se ha asumido que los componentes son independientes; si así no fuera, se debería incluir un término con la covarianza correspondiente.

En la práctica, se considera que U , E y P se distribuyen de forma normal, la varianza de U y de E debe estimarse por mediciones directas, la de P se obtiene mediante la calibración del termómetro patrón, usualmente en un laboratorio acreditado especialmente para ello, y para la varianza de D se asume que el error que introduce la división de escala se distribuye de mane-

ra uniforme entre $T_m - d/2$ y $T_m + d/2$, donde d es la división de escala¹.

Cálculo de la incertidumbre mediante el bootstrap

El *bootstrap* es un método para la estimación de varianzas desarrollado por Bradley Efron^{7,8}.

1) Se parte de una serie de n determinaciones que, en nuestro caso, asumiremos realizadas en diferentes puntos de la estufa elegidos al azar, y obtenemos la media muestral.

2) En las n temperaturas obtenidas se realiza un muestreo con repetición de tamaño n , de forma que algunas temperaturas pueden no ser seleccionadas ninguna vez y otras ser seleccionadas 2, 3 o más veces. Llamaremos *muestra bootstrap* a la muestra obtenida por este procedimiento, a continuación calculamos su media (la llamaremos *media bootstrap*).

3) Repetimos el paso 2 un número B de veces (se recomienda que sea al menos 200), con ello obtenemos B medias *bootstrap*.

4) Se calcula el error estándar de la temperatura como: $[\Sigma(\text{media bootstrap} - \text{media muestral})^2/B]^{1/2}$.

Una comparación de este procedimiento con el método delta muestra que el principal inconveniente del *bootstrap* es el esfuerzo de cálculo requerido para realizar al menos 200 remuestreos, aunque este esfuerzo ha quedado reducido al incluirse el *bootstrap* como instrucción estándar en algunos programas. A cambio, no requiere ninguna de las asunciones realizadas en el método delta: no es necesario conocer los factores que influyen en la temperatura medida, no se asume ningún tipo de relación específica entre estos factores y la temperatura y no se asume independencia.

Procedimiento de simulación

Para comparar directamente el *bootstrap* y el método delta se ha realizado una simulación por ordenador con el siguiente método:

1) Se generan al azar 40 mediciones de temperatura en una estufa que se supone regulada a 37 °C, con uniformidad = 0,75, estabilidad = 0,30, incertidumbre del patrón = 0,12, y división de escala del termómetro = 0,1.

2) Se realiza un *bootstrap* de tamaño $B = 1.000$ sobre las 40 mediciones obtenidas y se calcula su error estándar (lo denotamos por *EEB*) y, por simplicidad, asumiendo normalidad se calcula el correspondiente intervalo de confianza al 95%: $37 \pm 1,96EEB$.

3) En cuatro puntos diferentes se generan al azar diez temperaturas en cada uno, de forma que las cuatro mediciones de orden i (esto es, las cuatro temperaturas tomadas en diferentes puntos) se consideran simultáneas y tienen distribución normal de media 37

y varianza σ_u^2 (porque están afectadas por la uniformidad pero no por la estabilidad). Las diez mediciones de cada punto son normales de varianza σ_e^2 (porque están afectadas por la estabilidad pero no por la uniformidad). Estas 40 determinaciones se emplean para estimar el error estándar de la temperatura por el método delta y el correspondiente intervalo de confianza: $37 \pm 1,96EED$.

4) Se generan al azar 40 temperaturas con las mismas condiciones de uniformidad y estabilidad anteriores, se calcula su media y se comprueba si está incluida en los intervalos de confianza del *bootstrap* y del delta.

5) Se repite el punto 4 en 100 ocasiones con lo que se obtiene el porcentaje de medias que quedan incluidas dentro de cada intervalo de confianza. Llamaremos *porcentaje de cobertura* al porcentaje obtenido por este procedimiento.

Este procedimiento se repitió 1.000 veces para obtener la distribución del porcentaje de cobertura obtenido en los intervalos de confianza del *bootstrap* y del delta. De esta manera se pretende saber si el porcentaje de cobertura realmente obtenido se acerca al 95% nominal del intervalo de confianza.

Una vez realizada la simulación anterior se introdujeron algunas modificaciones en las condiciones iniciales:

1) En las simulaciones segunda y tercera se consideró el caso de una variable que influya en la temperatura pero cuya existencia es desconocida para la persona que realiza la calibración de la estufa. Se analizó su influencia en los porcentajes de cobertura de cada intervalo de confianza considerando que tenía, sucesivamente, varianzas de 0,1 y 0,2.

2) En las simulaciones cuarta y quinta se consideró la posibilidad de que la uniformidad y la estabilidad no fueran independientes sino que tuvieran, respectivamente, un coeficiente de correlación de 0,2 y -0,2. Este rango es arbitrario y pretende mostrar el efecto sobre las mediciones de una correlación relativamente baja.

Resultados

La **tabla 1** muestra los porcentajes de cobertura medios obtenidos con cada intervalo de confianza en cada simulación. Con la única excepción de la simulación cuarta, los porcentajes de cobertura medios son más próximos a 95% en el *bootstrap* que en el método delta. La **figura 1** muestra la distribución del porcentaje de cobertura en las simulaciones 1 y 2. Puede observarse cómo el *bootstrap* consigue resultados en el entorno del 95% un gran número de veces, mientras que el delta tiene un 9% de sus resultados con un 100% de cobertura (simulación 1) y una cola izquierda más grande que la del *bootstrap*. Cuando aparece un componente desconocido en la temperatura (simulación 2), el *bootstrap* mantiene la misma distribución del porcentaje de co-

Tabla 1. Media de los errores estándar y media de los porcentajes de cobertura en los intervalos de confianza obtenidos por los métodos *bootstrap* y *delta*

Simulación	EEB	EED	Porcentaje de cobertura	
			<i>Bootstrap</i>	<i>Delta</i>
1	0,0286	0,0307	94,10	93,67
2	0,0305	0,0303	93,80	92,30
3	0,0330	0,0304	93,76	91,32
4	0,0256	0,0309	94,15	94,91
5	0,0310	0,0302	93,74	92,10

EEB: error estándar método *bootstrap*.
EED: error estándar método *delta*.

bertura, mientras que el *delta* se desplaza hacia valores inferiores. Este efecto es aún más marcado en la simulación 3 (no se muestra en las figuras).

La **figura 2** muestra los resultados de las simulaciones 4 y 5 (que corresponden a la ausencia de independencia entre la uniformidad y la estabilidad). De nuevo el *bootstrap* consigue un gran número de coberturas cercanas al 95% mientras que el *delta* tiene una distribución muy variable que se desplaza hacia la izquierda al aumentar el coeficiente de correlación.

Figura 1. Distribución del porcentaje de cobertura obtenido por los métodos *delta* y *bootstrap* en las simulaciones 1 (figura 1a**) y 2 (**figura 1b**)**

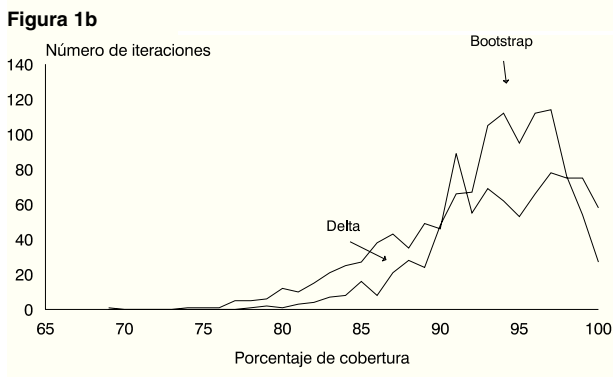
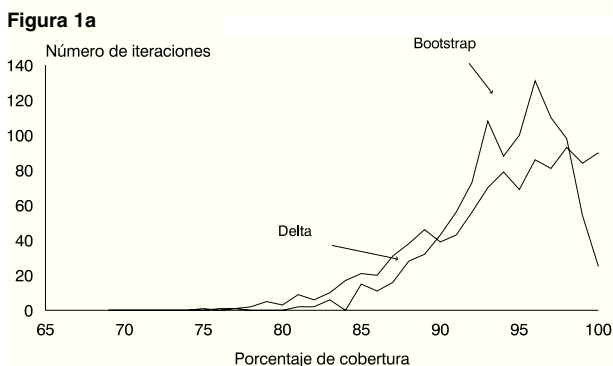
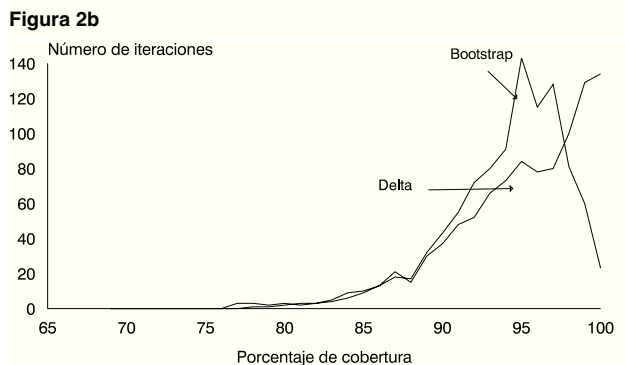
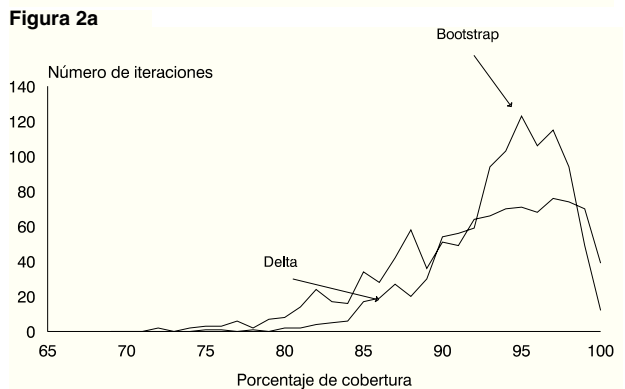


Figura 2. Distribución del porcentaje de cobertura obtenido por los métodos *delta* y *bootstrap* en las simulaciones 4 (figura 2a**) y 5 (**figura 2b**)**



Finalmente, se han querido mostrar en la misma figura los resultados de todas las simulaciones para hacer una comparación global (**figura 3**). Las cinco distribuciones correspondientes al *bootstrap* son casi superponibles: el *bootstrap* obtiene la misma distribución del porcentaje de cobertura con independencia de las condiciones consideradas. En cambio, el *delta* muestra mayor irregularidad: en unas ocasiones la moda llega a ser el 100% de cobertura y en otras está por debajo del 95%. En dos aspectos el *delta* muestra una gran regularidad: la cola izquierda es mayor que en el *bootstrap* (un número grande de veces se obtienen porcentajes de cobertura inferiores al 90%) y el número de ocasiones en que el *delta* obtiene porcentajes próximos al nominal (95%) es siempre menor que en el *bootstrap*.

Discusión

Ninguna medición puede determinar exactamente el valor de la magnitud que se pretende medir; por ello, es necesario incluir algún tipo de medida relativa al rango de valores plausibles que esa magnitud puede tomar. Esta práctica ya es habitual en la investigación médi-

Figura 3. Distribución de los porcentajes de cobertura obtenidos por los métodos delta (figura 3a) y bootstrap (figura 3b) en todas las simulaciones

Figura 3a

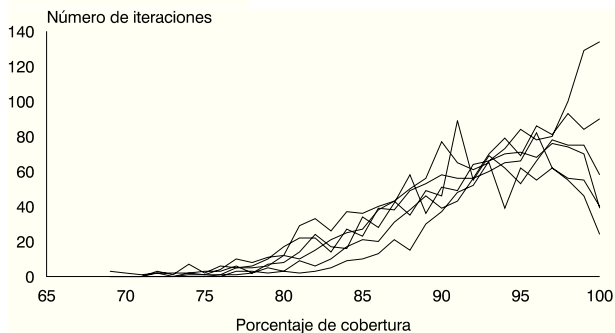
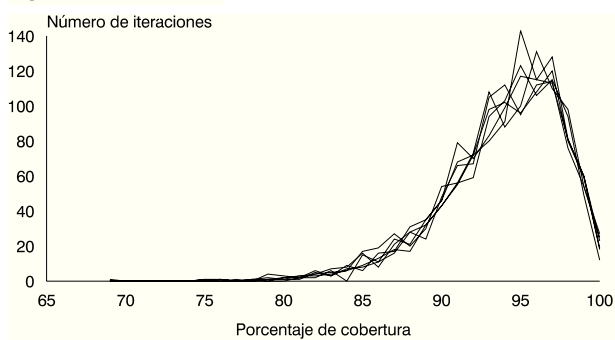


Figura 3b



ca, donde se calculan frecuentemente los intervalos de confianza de incidencias, riesgos relativos y de otros indicadores. Los procesos de acreditación de laboratorios clínicos y de salud pública amplían el territorio del cálculo de incertidumbres a la práctica clínica habitual; es decir, permitirán el cálculo de intervalos de confianza para el resultado clínico individual. Para ello es necesario unificar criterios sobre cómo deben calcularse estos intervalos de confianza y sobre su significado. La guía ISO establece algunos criterios para ello.

Algunas limitaciones de la guía ISO han sido señaladas; en particular, Gleser ha llamado la atención sobre que la guía ISO emplea métodos frecuentistas para el cálculo de incertidumbres y, sin embargo, realiza una interpretación bayesiana del resultado («... caracteriza el intervalo de valores, dentro del cual se encuentra el valor verdadero de la cantidad medida»); según sus propias palabras, esta interpretación carece de base científica⁹ por lo que ha sugerido interpretaciones alternativas.

La simulación realizada en este trabajo muestra que el método delta (propuesto en la guía ISO para el cálculo de incertidumbres)¹ tiene porcentajes medios de cobertura similares (algo inferiores, en general) a los obtenidos por el *bootstrap*; sin embargo, la distribución de estos porcentajes de cobertura tiene una dispersión

mucho mayor en el delta: el *bootstrap* tiene coberturas habitualmente próximas al nominal del 95%, mientras que el delta presenta una frecuencia alta de coberturas del 100% y por debajo del 95%.

El empleo del *bootstrap* no requiere asunciones estadísticas previas, mientras que el empleo del método delta requiere asumir que se conocen todas las variables que influyen en la magnitud medida y que se conoce toda la matriz de covarianzas entre ellas. De hecho, el *bootstrap* ha ofrecido resultados muy similares cuando se ha incluido una variable desconocida (simulaciones 2 y 3) o cuando las variables no eran independientes entre sí (simulaciones 4 y 5), mientras que el método delta ha empeorado sus resultados en ambas situaciones.

En la práctica, esto debe significar que el método delta debe limitarse a la medición de magnitudes que son definidas en función de otras variables, y no para la medición de aquellas magnitudes que disponen de instrumentos de medida propios. Por ejemplo, supóngase que un laboratorio desea conocer la incertidumbre de la concentración de una solución de ClNa en agua, medida en peso/volumen. La concentración no se mide directamente sino que se calcula como el resultado de dividir el peso de la sal empleada entre el volumen de agua. En este caso, emplear el método delta sería adecuado. La temperatura de una estufa, en cambio, es una variable primaria; es decir, es medida directamente. La función que la hace depender de la uniformidad y la estabilidad de la estufa no es más que una suposición. Aplicar el método delta en este caso puede conducir a graves errores por mala especificación del modelo, bien por desconocimiento de una covarianza entre ellas, bien por desconocimiento de otras variables que intervienen en la temperatura interna de la estufa (por ejemplo, la temperatura del laboratorio). Como hemos mostrado en nuestra simulación, el método delta no es capaz de detectar estos errores, mientras que el *bootstrap* es un método de estimación robusto porque obtiene resultados de cobertura similares aunque las condiciones varíen.

El principal problema relacionado con el uso del *bootstrap* es la capacidad de cálculo que requiere. Esta podría ser una dificultad insalvable en la era preinformática, pero no en el presente. Cada una de las simulaciones realizadas para este trabajo requería el cálculo de 1.000 estimaciones por *bootstrap*, con 1.000 iteraciones cada una. Cada simulación se llevó a cabo en catorce segundos mediante un programa *ad hoc* elaborado en TurboPascal 6.0¹⁰. En la actualidad algunos paquetes estadísticos incorporan el *bootstrap*.

En conclusión, el *bootstrap* permite un cálculo de incertidumbres que, en nuestra simulación, resulta más fiable que el método delta. El método delta debería restringirse a las situaciones en que la magnitud considerada se obtiene como resultado de la medición de otras magnitudes.

Bibliografía

1. ISO Technical Advisory Group, Working Group 3. Guide to the Expression of Uncertainty in Measurement. Ginebra: International Organization for Standardization; 1993.
 2. Guía para la expresión de la incertidumbre de medida en las calibraciones. Traducción del documento WECC-19-1990. Madrid: Entidad Nacional de Acreditación; 1991.
 3. EAL. Expression of the Uncertainty of Measurement in Calibration. European cooperation for Accreditation of Laboratories, 1997.
 4. ENAC. Expresión de la incertidumbre de la medida en calibración. Documento CEA— ENAC-LC/02. Traducción de la referencia 3. Madrid: Entidad Nacional de Acreditación; 1998.
 5. Spiegel MR, Abellanas L. Fórmulas y tablas de matemática aplicada. Madrid: McGraw-Hill; 1992. p. 84-8.
 6. Cramer H. Mathematical methods of statistics. Princeton: Princeton University Press; 1946. p. 359.
 7. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1986;1:54-77.
 8. Efron B. Nonparametric estimates of standard error: the jackknife, the bootstrap, and other resampling plans. *Biometrika* 1981; 68:589-99.
 9. Gleser LJ. Assessing Uncertainty in Measurement. *Statistical Science* 1998;13:277-90.
 10. Turbo Pascal 6.0 programmer's guide. Borland International, Scots Valley; 1990.
-
-