

EVALUACIÓN DE RESULTADOS DE ENSAYOS DE APTITUD

J. Laso Sánchez¹ y A. Peris García-Patrón¹

¹Gabinete de Servicios para la Calidad S.A.L., C/ Caridad 32, 28007 Madrid; e-mail: gscsal@gscsal.com

RESUMEN. El objeto de este trabajo es reflexionar sobre los diferentes sistemas empleados para la evaluación de los laboratorios participantes en ensayos de aptitud, que han sido desarrollados en los últimos años, profundizando sobre los puntos débiles y fuertes de cada uno de ellos, de modo que permita a los laboratorios usuarios de los servicios de proveedores ensayos de intercomparación, disponer de una herramienta adicional para la toma de decisiones y realización de acciones.

1.- Introducción

Los ensayos de intercomparación son una de las pruebas de la calidad de las medidas analíticas, que permiten a los laboratorios asegurar la comparabilidad de sus resultados y, en última instancia, conocer mejor el comportamiento de sus métodos, respecto al conjunto de participantes y así, mejorar y corregir dicho comportamiento, cuando sea necesario.

Como medida de la comparabilidad analítica y “termómetro” de la calidad en la aplicación de un método, es usado como herramienta de la evaluación de la calidad de los resultados de los ensayos, tal como solicita la Norma de referencia de valoración de la competencia técnica de laboratorios de ensayo UNE-EN ISO/IEC 17025 y, como requisito que debe implantarse, establecido en España por la Entidad Nacional de Acreditación, a través de sus Criterios Generales de Acreditación, Notas técnicas y Guías emitidas al respecto (1, 2, 3).

En general los ensayos de aptitud son realizados por un proveedor, que distribuye entre los diferentes laboratorios, una muestra que debería ser idéntica, por tanto homogénea y estable, para que los participantes le remitan los resultados obtenidos en el análisis de la misma, de acuerdo a unas instrucciones pactadas, y el proveedor realice una evaluación que informe a cada laboratorio sobre la adecuada realización de los análisis remitidos.

Para realizar esta evaluación en el caso de los análisis cuantitativos se han utilizado una serie de parámetros descritos en general en la norma ISO 43, que han sido complementados y desarrollados con más grado de detalle en otros documentos tales como el documento ISO 13528, ó el protocolo IUPAC.

2.- Sistema de evaluación y sus características

Los sistemas de evaluación varían en cuanto al parámetro calculado para evaluar el desempeño de un laboratorio y su forma de obtención.

2.1 Puntuación Z

El sistema más empleado en la actualidad es la denominada puntuación Z.

Ésta se obtiene de acuerdo a la fórmula:

$$z = \frac{V_L - V_A}{\sigma} \quad (1)$$

Donde:

VL es cada uno de los datos aportados por los laboratorios participantes.

VA es el valor asignado por el evaluador a la muestra analizada, para el parámetro objeto de estudio.

σ : es el parámetro que caracteriza al intervalo construido entorno al valor asignado como más probable o verdadero.

(VA), dentro del cual debe situarse el resultado aportado por cada laboratorio, para ser considerado como “adecuado”, con una probabilidad dada.

En su uso, subyacen las siguientes suposiciones:

-Los resultados de los laboratorios siguen una distribución normal.

-El valor central de dicha distribución es el valor asignado (VA).

-La variabilidad depende de σ .

Como al conjunto de resultados “adecuados” de los laboratorios se le atribuye un comportamiento normal y, habitualmente el intervalo de aceptación se construye con un nivel de confianza del 95%, la interpretación de los resultados del z-score es:

- $|Z| \leq 2$ Satisfactorio: Se produce en el 95% de los casos, si los resultados son correctos.

- $2 < |Z| \leq 3$ Sospechoso: Se produce en el 5% de los casos si los resultados son correctos.

- $Z > 3$ Incorrecto

Por lo tanto lo que suponemos es que nuestro laboratorio para ser considerado aceptable, se comporta como si perteneciese a una población de valor más probable VA y de variabilidad sigma.

-Nota: Si nuestro laboratorio, ó nuestro método, tiene una variabilidad mucho mejor ó peor que la sigma utilizada, obtendremos resultados que no, evaluarán nuestra situación real, ya que nos compararan contra un colectivo al que no pertenecemos.

El segundo problema importante es la forma en que se determinan VA y sigma, que ha sido desarrollado de manera prolíja, por los diferentes documentos citados anteriormente.

2.1.1 Valor asignado VA

El Valor asignado puede ser obtenido por diferentes planteamientos que deberían garantizar la seguridad del mismo y su adecuación a todos los participantes del ejercicio.

Los sistemas propuestos se resumen a continuación.

2.1.1.1 Utilización de resultado obtenido por un laboratorio de referencia

Requiere evidentemente una garantía absoluta del Valor del laboratorio así como de la incertidumbre con la que ha obtenido el resultado, de modo que ésta sea compatible con la sigma que se utiliza. Para ello no debería ser superior a 0,3 del Valor de sigma para no producir efectos significativos (menores del 5%) en la evaluación de los resultados.

2.1.1.2 Utilización de material de referencia certificado

La utilización de un material de referencia certificado permite la asignación de un Valor seguro siempre que la incertidumbre del material cumpla los criterios indicados anteriormente y su Valor sea independiente del método utilizado para obtenerlo. En caso contrario es necesario realizar un aviso a los participantes para indicarles que no existe una seguridad en su evaluación, si utilizan un método diferente del utilizado para la certificación del material.

2.1.1.3 Comparación contra un material de referencia certificado

En la práctica se utiliza un único laboratorio para establecer el Valor asignado mediante la comparación de la muestra objeto del estudio con un material de referencia. Es una mezcla de los dos sistemas anteriormente indicados, y por tanto requiere garantizar la incertidumbre del laboratorio que realiza la comparación, la incertidumbre del material de referencia utilizado, y la adecuación del método utilizado.

2.1.1.4 Formulación

El material objeto de ensayo se prepara mediante mezcla de constituyentes en proporciones especificadas, ó adicionando una cantidad conocida del constituyente sobre

un material base. Una de las aplicaciones típicas es la preparación de muestras adicionadas (spiked simples).

El valor asignado se obtiene mediante cálculos a partir de las proporciones utilizadas y por tanto su incertidumbre (para la que se debe evaluar su adecuación a la sigma), mediante una combinación de incertidumbres siguiendo los principios de la Guía de expresión de incertidumbres de medidas (7).

Para el caso de método de análisis químico deberían tenerse en cuenta algunos aspectos, de cara a garantizar que la incertidumbre ha sido bien calculada:

1. El material base está libre del constituyente añadido ó en una proporción tal que no afecta al resultado de la adición.
2. Se garantiza una mezcla homogénea de todos los constituyentes.
3. No existe interacción entre los constituyentes y la matriz.
4. Se garantiza una estabilidad de la adición.

2.1.1.5 Por consenso a partir de laboratorios expertos

En este caso se obtiene el Valor asignado a partir de los datos obtenidos en una experimento previo. ó paralelo al ensayo de aptitud, en el que se envían porciones de la muestra, a diferentes laboratorios considerados expertos.

Requiere de una selección cuidadosa de los laboratorios expertos, así como de la incertidumbre con la que estos, informan de los resultados.

Supone disponer de un sistema adecuado basado en estadística robusta (que minimice el efecto de los resultados anómalos) para obtener el valor asignado.

La norma ISO 13528 plantea un sistema de cálculo a partir del algoritmo A.

Plantean dificultades operativas ya que el número de laboratorios expertos estará limitado, por lo que los datos pueden no tener significación estadística suficiente.

2.1.1.6 Valor del consenso obtenido de los participantes

El resultado se obtiene a partir de cálculos estadísticos, obtenidos del conjunto de datos suministrados por los participantes.

Los problemas inherentes a este sistema incluyen:

- 1) La existencia de resultados anómalos y su tratamiento, a través de la utilización de test estadísticos de eliminación.
- 2) La utilización de sistemas de obtención de un Valor asignado que no se vean influenciados por la presencia de datos anómalos, bien a través de su eliminación previa, bien mediante la utilización de parámetros estadísticos robustos como por ejemplo media robusta, mediana, ó moda.
- 3) La evaluación de ausencias de sesgos entre los diferentes métodos utilizados por los participantes, lo que puede dar lugar a distribuciones plurimodales, que impiden la obtención de un parámetro representativo para el valor asignado.

Sin embargo este sistema lo mismo que el e) permite evaluar a laboratorios realizando análisis sobre muestras naturales, en las que el analito de interés, se encuentra presente de forma natural y requiere la comprobación de que el valor asignado se obtiene con una incertidumbre adecuada al SIGMA de evaluación.

2.1.2 Valor de variabilidad (SIGMA)

Al igual que para el caso del valor asignado, los documentos establecen diversos sistemas para su establecimiento si bien se plantean dos formas generales de obtención:

a) A partir de información previa al ejercicio en base a la suposición de su aplicabilidad a los datos obtenidos. Clásicamente se denomina sistema de adecuación al uso o sigma objetivo. En este caso tal como indica el protocolo IUPAC no se evalúa, cómo están trabajando los laboratorios, sino cómo deberían trabajar, en función de suposiciones previas.

Su problema fundamental consiste en garantizar que el Valor adecuado para el uso previsto para el SIGMA, es correcto, ya que en muchos casos legales, este valor se deriva de límites que no provienen de ningún valor experimental y pueden realizar evaluaciones no satisfactorias.

Por ello en la práctica se complementan con pruebas adicionales de comparación de la sigma objetivo con la sigma obtenida por los participantes durante el ejercicio ó la desviación estándar de reproducibilidad del método utilizado, si este es único y normalizado, para valorar que no existen diferencias estadísticamente significativas.

b) A partir de los datos experimentales obtenidos durante el ejercicio una vez estudiados estadísticamente para garantizar su fiabilidad y eliminando el efecto de anómalos.

De acuerdo a estos dos planteamientos se estructuran las siguientes prácticas:

2.1.2.1 Valor por prescripción

Se establece una desviación estándar en base a los datos requeridos para su interpretación ó a los requisitos establecidos en legislación.

2.1.2.2 Por percepción

Se establece la desviación estándar en base a las suposiciones realizadas por el proveedor ó coordinador, según considere qué debe ser aceptable. Se utiliza en muchos casos en circuitos, organizados por entes con capacidad legal, que quieren evaluar la capacidad de los laboratorios que dependen de ellos, por ejemplo la incertidumbre que desean que tengan los laboratorios.

2.1.2.3 En base a un método general

El valor de la desviación estándar se obtiene de un modelo general de reproducibilidad del método de medida. Existen diferentes modelos utilizados como por ejemplo el de

Horwitz. Es necesario garantizar una adecuación suficiente al uso.

2.1.2.4 De los resultados de un experimento de precisión

Cuando el método utilizado por todos los laboratorios es el mismo y existe información sobre repetibilidad y reproducibilidad del método, la desviación estándar utilizada como sigma objetivo puede obtenerse, a partir de los datos de reproducibilidad del método normalizado.

Requiere por tanto, que todos los laboratorios utilicen el mismo método, y que los experimentos iniciales realizados para su publicación como método normalizado, y la obtención de los parámetros representativos, sean suficientemente extensos, en cuanto al número de participantes, tipos de muestras, y niveles de analito.

2.1.2.5 A partir de los resultados obtenidos en el ejercicio de intercomparación.

Suponen realizar los cálculos para obtener la desviación estándar entre los participantes.

Requieren la utilización de test estadísticos, que eliminen los resultados anómalos que pueden estar presentes, ó de estadísticas robustas, ya que la influencia de los resultados anómalos, es muy importante en el caso de la sigma: mucho más que en el caso del Valor asignado.

Suponen la existencia de diferencias entre distintas rondas de un mismo ejercicio, que pueden dificultar la forma en que se realizará una evaluación, aunque existen sistemáticas para realizarla.

3.- Evaluaciones en base a la incertidumbre

Las normas de referencia, plantean la posibilidad de realización de otros tipos de evaluaciones, basadas en la comparación de las diferencias de los resultados frente al Valor asignado, y las incertidumbres del mismo, con el valor remitido por el laboratorio y la incertidumbre del mismo.

En este apartado se podrían citar:

3.1 El número En

Compara la diferencia del valor del laboratorio y del Valor asignado, con las incertidumbres expandidas procedentes del ejercicio y del laboratorio, según la siguiente fórmula:

$$E_n = \frac{|V_L - V_A|}{\sqrt{U_L^2 + U_A^2}} \quad (2)$$

En este caso los criterios de aceptación son que el número En sea menor de uno, si bien se podría admitir que en el 5% de los casos, estuviese comprendido entre 1 y 1,5.

3.2 Las puntuaciones zeta

Comparan la diferencia del valor del laboratorio y su valor asignado, con las incertidumbres combinadas procedentes del ejercicio y del laboratorio, según la siguiente fórmula:

$$zeta = \frac{|V_L - V_A|}{\sqrt{U_L^2 + U_A^2}} \quad (3)$$

La interpretación de la puntuación zeta es la misma que era la de la puntuación Z.

3.3 Otros tipos de puntuaciones que combinan la sigma objetivo del ejercicio con la del propio laboratorio.

La característica de estas evaluaciones, en el caso de a) y b), es que son independientes de la sigma objetivo, ya que lo que se evalúa es la declaración que hace el propio laboratorio, sobre la seguridad de sus resultados, y por tanto la veracidad de sus afirmaciones. Requiere a cambio que cada laboratorio incluya la incertidumbre que declara en su resultado.

Esto permite que los laboratorios participen en ensayos de aptitud en los que el conjunto no utiliza su método, ó no tiene sus prestaciones, ya sea porque son mejores ó peores. Además suministra información sobre la correcta evaluación de la incertidumbre realizada por el laboratorio.

4.- Evaluaciones realizadas en los límites analíticos

En algunos casos, el contenido del analito en las muestras naturales, que realiza habitualmente el laboratorio, se encuentran en zonas próximas a los límites de cuantificación de los resultados. Este hecho está produciendo problemas en las evaluaciones, debido a que en muchos casos los proveedores no evalúan los resultados, que no son declarados como cuantitativos, es decir se indican menor de un límite. Por ello en algunos casos se solicita por parte del proveedor, que independientemente de los límites de cuantificación se indique un resultado, que se encuentra en un rango que el laboratorio no ha comprobado. De este modo se obtienen un conjunto de resultados que se pueden procesar estadísticamente, y que se pueden valorar de acuerdo a los criterios indicados anteriormente.

Sin embargo si el ensayo de aptitud trata de reproducir la realidad vivida por el laboratorio en su práctica cotidiana, no es aconsejable el envío de resultados por debajo de los límites que el laboratorio tiene comprobar ya que, independientemente de la evaluación obtenida, no se pueden tomar ningún tipo de acciones, puesto que si la evaluación es positiva, esto no garantiza que el método funcione en ese rango ya que el número de pruebas es limitado, y si la evaluación es negativa, tampoco se deben tomar acciones, ya que el laboratorio tiene declarado que no garantiza que sus resultados sean satisfactorios en ese punto.

A continuación se propone una propuesta de evaluación utilizada por GSC en sus ejercicios.

La propuesta se basa en:

-Debe establecerse un Valor asignado independiente del conjunto de valores obtenidos en el ejercicio puesto que no será posible realizar cálculos estadísticos con valores no numéricos. Para ello se tendrán en cuenta los resultados obtenidos en los análisis previos de la muestra, los límites de cuantificación habituales en el sector o prescritos en la normativa.

-Debe establecerse una sigma objetivo ó un límite de variabilidad basado en el límite de detección existente.

-A partir de estos datos se establecen unos intervalos que permiten la valoración de un laboratorio que ha dado sus datos como resultados cualitativos.

Si $LQ \leq V_A$ (Correcto)

Si $LQ \geq V_A - 2\sigma$ (Correcto)

Si $V_A - 2\sigma > LQ \geq V_A - 3\sigma$ (Cuestionable)

Si $V_A - 3\sigma > LQ$ (No satisfactorio)

5.- Ejemplo de evaluación

A continuación se plantean algunos ejemplos de evaluación que demuestran la importancia de la selección de las fuentes en las que se basan los criterios para la evaluación de los resultados de modo que puedan ser útiles para los laboratorios.

5.1 Conductividad en aguas continentales

El laboratorio participa en 2 circuitos de aptitud incluyendo la determinación del parámetro de conductividad a 20° C.

Ambos ejercicios utilizan diferentes sistemas de evaluación: en el primer caso se utiliza un Valor asignado obtenido como media robusta y una sigma objetivo procedente de la reglamentación; en el segundo caso se utiliza una media de los laboratorios una vez eliminados anómalos, a través de un test de estadística robusta, obteniendo asimismo de ese conjunto de resultados, la desviación estándar de consenso.

Debido a la realización de diferentes rondas en los ejercicios se produce la coincidencia de dos participaciones en un plazo corto de tiempo.

Los resultados obtenidos por el laboratorio son:

$V_L = 878$ $Z = 4,05$ (Sigma consenso)

$V_L = 1045$ $Z = 1,02$ (Sigma reglamentado)

Como consecuencia de dichos resultados se realiza un estudio de las causas para lo que se examinan los criterios de evaluación de los dos ejercicios, fundamentalmente las σ utilizadas por los mismos que resultan ser:

- σ consenso = 21,2

- σ reglamentado = 104,5

-Nota: La causa de las discrepancias en las dos valoraciones es debida fundamentalmente a la utilización de sigmas de evaluación muy diferentes.

Una vez revisado el ensayo se comprueba que se ha producido un ajuste inadecuado del control de temperatura

de la conductividad que se encuentra fijada a 25° C, lo que produce como consecuencia un sesgo de cara a la conductividad expresada a 20° C de aproximadamente 10%. Este problema fue detectado por el sistema de sigma de consenso.

Sin embargo, la sigma de consenso puede ser, en ocasiones y, debido a la “casualidad” estadística, muy restrictiva, lo que tampoco garantiza una valoración justa de los resultados en todos los casos. Así, por ejemplo, en un ensayo que afecta al mismo parámetro, conductividad a 20°C, en aguas de consumo, se obtuvieron los resultados experimentales:

-VAsignada=392 $\mu\text{S}/\text{cm}$, conseguida con una σ consenso=6,6. Como consecuencia de estos resultados, un laboratorio que informaba un resultado como $V_L=420$, de usar directamente el valor de sigma de consenso, obtendrá una puntuación $z=4.2$.

En este caso, el organizador, disponía de valores de consenso históricos que mostraban que, para este parámetro, la sigma ponderada, consecuencia de múltiples ensayos de intercomparación, de este parámetro, y en este tipo de muestra, se situaba en valores próximos a un 4% del valor obtenido; el proveedor evaluó que el valor de sigma obtenido (inferior al 2%, lo que no cubre apenas diferencias de 1°C, como incertidumbre del ajuste de la temperatura de referencia) no correspondía con el sigma objetivo conocido, en base a históricos; por ello, la evaluación del proveedor fue realizada con los valores σ histórica ponderada, considerada como la esperada si se conoce el comportamiento experimental del parámetro, en este caso, correspondiente al valor de 16.3, por lo que el mencionado laboratorio obtuvo resultado de $z\text{-score}=1.7$, claramente satisfactorio.

6.- Conclusiones

1. Existen diferentes tratamientos de los resultados que pueden alterar los datos de la evaluación, por lo que es recomendable que los laboratorios dispongan de herramientas internas para la evaluación de sus participaciones y comprendan la forma utilizada por los proveedores en su evaluación.

2. La evaluación con la puntuación z sólo indica si el laboratorio pertenece a una población de valor asignado VA, y de variabilidad σ .

Por lo que pueden existir puntuaciones no adecuadas dependiendo, de qué valores se asignen a dichos parámetros, y de que nuestros métodos sean adecuados y nuestras incertidumbres compatibles.

3. Existen diversos sistemas para la determinación de los parámetros VA y σ que utilizan:

- Valores externos y previos al ejercicio.
- Valores obtenidos en el propio ejercicio.

Los primeros garantizan una constancia, y una independencia de los resultados de los participantes, pero no siempre se basan en datos experimentales, lo que puede suscitar evaluaciones inútiles para detectar problemas, ya que no evalúan como se comportan los laboratorios, sino

cómo deberían comportarse.

Los segundos basados en datos experimentales, permiten una evaluación más realista del conjunto de laboratorios, pero exigen un tratamiento que disminuya la influencia de los anómalos en la obtención de VA y σ .

Nota: Ninguno de los dos sistemas (objetivo ó consenso) constituye por sí solo un sistema completamente seguro, para ser aplicados de manera automática, sin el estudio de los resultados obtenidos, antes de proceder a la evaluación.

Es imprescindible, en el caso de la definición de parámetros por consenso, la utilización de sistemas basados en estadísticas robustas, preferiblemente que eliminen anómalos antes de realizar los cálculos.

Este hecho es particularmente importante en el caso de la determinación de la σ .

1) Se debe solicitar a los proveedores de los ejercicios que elaboren sistemas de evaluación en las zonas límites, de modo que los laboratorios puedan presentar los resultados que realmente informarían, si analizaran la muestra de intercomparación como una muestra normal, sin ser eliminados de la evaluación a pesar de haber participado.

2) Dado que la incertidumbre:

a) Es un parámetro fundamental para interpretar los resultados.

b) Depende del método y equipos empleados, así como de la pericia del laboratorio.

c) Es declarada por éste en base a estimaciones de metodología definida.

d) Puede ser útil de cara a que el cliente seleccione un laboratorio para una aplicación específica.

Es aconsejable que las evaluaciones comiencen a tenerla en cuenta. Esto permitirá:

a) Valorar si la incertidumbre declarada esta correctamente estimada.

b) Valorar si el laboratorio informa correctamente, independientemente de que utilice un método de variabilidad (σ) más alta que la σ de evaluación.

c) Valorar realmente la capacidad de los laboratorios que por sus características especiales tienen variabilidades muy inferiores (σ pequeñas) al conjunto y por tanto obtienen en general resultados excelentes.

Referencias

- CGA-ENAC-LE Rev. 4 “Criterios generales para la acreditación de laboratorios de ensayo y calibración según norma UNE EN ISO/IEC 17025”
- NT-03 Rev. 3 “Política de ENAC sobre Intercomparaciones”.
- G-ENAC-14 Rev. 1 “Guía sobre la participación en programas de intercomparación”
- UNE 66543-IIN (ISO Guide 43-1) “Ensayos de aptitud por intercomparación de laboratorios. Parte 1: Desarrollo y aplicación de programas de ensayos de aptitud”
- ISO 13528:2005 “Statistical methods for use in proficiency testing by interlaboratory comparisons”.
- Protocolo IUPAC Technical Report “The international harmonized protocol for the proficiency testing of analytical chemistry laboratories”. *Pure Appl.Chem.*, Vol 78, nº 1, pp.145-196, 2006. M. Thompson, S. L.R. Ellison y R. Wood.
- Guía GUM “Guía para la expresión de la incertidumbre de medida”