GUÍA DE VALIDACIÓN DE MÉTODOS ANALÍTICOS

INTRODUCCIÓN 3

ALCANCE 4

1. LINEALIDAD DE LA CURVA DE CALIBRACIÓN 5

NOTA SOBRE EL COEFICIENTE DE CORRELACIÓN COMO PRUEBA DE LINEALIDAD 7

NOTA SOBRE CURVAS DE CALIBRACIÓN QUE PASAN POR EL ORIGEN 9

TEST DE CARENCIA DE AJUSTE. GUÍA ISO 11095 10

TEST DE MANDEL ISO 8466-1 14

ALGUNAS CONSIDERACIONES PARA INTERPRETAR LOS TESTS DE LINEALIDAD 17

MÉTODO DE CALIBRACIÓN CON ADICIÓN DE ESTÁNDAR INTERNO 19

NOTA PARA MÉTODOS CROMATOGRÁFICOS ACOPLADOS A ESPECTROMETRÍA DE MASAS 20

NOTA PARA MÉTODOS ICP-OES E ICP-MS. 20

NOTA SOBRE CALIBRACIÓN EN MATRIZ (MATRIX-MATCHED CALIBRATION) 20

2. PRECISIÓN 21

ESTIMACIÓN DE LA PRECISIÓN DE ACUERDO A LA GUÍA ISO 5725 PARTE 3 21

ESTIMACIÓN DE LA PRECISIÓN DE REPETIBILIDAD sr MEDIANTE ANÁLISIS DE DUPLICADOS DE QAQC. 25

ESTIMACIÓN DE LA MÁXIMA DIFERENCIA TOLERABLE ENTRE DUPLICADOS 28

ALGUNAS OBSERVACIONES DE LOS ESTUDIOS DE PRECISIÓN 29

3. INCERTIDUMBRE DE MEDICIÓN 30

CÁLCULO DE INCERTIDUMBRE BASADO EN LA GUÍA GUM 32

ESPECIFICACIÓN DEL MODELO 32

IDENTIFICACIÓN DE LAS FUENTES DE INCERTIDUMBRE 33

CUANTIFICACIÓN DE LAS INCERTIDUMBRES ESTÁNDAR DE LAS FUENTES DE ERROR 33

COMBINACIÓN DE LAS INCERTIDUMBRES ESTÁNDAR 34

CÁLCULO DE INCERTIDUMBRE EXPANDIDA 35

EJEMPLO DE CÁLCULO DE INCERTIDUMBRE MÉTODO ISO-GUM. MÉTODO DE DETERMINACIÓN DE HG EN CHORITOS POR ICP-MS. 36

DISEÑO EXPERIMENTAL DE LA CURVA PARA MINIMIZAR LA INCERTIDUMBRE DE CALIBRACIÓN 39

4. LÍMITES DE DETECCIÓN Y CUANTIFICACIÓN 46
5. SESGO Y RECUPERACIÓN 48
EXISTENCIA DE UN MATERIAL DE REFERENCIA CERTIFICADO (MRC) 48
OBSERVACIONES E INTERPRETACIÓN DEL TEST ESTADÍSTICO 49
RECUPERACIÓN: USO DE MUESTRA FORTIFICADA 51
OBSERVACIONES A LOS ESTUDIOS DE RECUPERACIÓN: 52
6. ROBUSTEZ 53
ESQUEMA GENERAL DE UN DISEÑO EXPERIMENTAL DE ROBUSTEZ PLACKETT-BURMAN 54
EJEMPLO DE ESTUDIO DE ROBUSTEZ. DETERMINACIÓN DE CLORANFENICOL EN SALMÓN 54
ALGUNAS CONSIDERACIONES DE LOS ESTUDIOS DE ROBUSTEZ 57
7. ESPECIFICIDAD/SELECTIVIDAD 58
8. REVALIDACIÓN Y/O VERIFICACIÓN DE UN MÉTODO ANALÍTICO DE ACUERDO A ESTA GUÍA. 61
BIBLIOGRAFÍA 63
CHECKLIST MÍNIMO 71

INTRODUCCIÓN

La validación y/o verificación de métodos analíticos es una actividad fundamental y no opcional para confirmar experimentalmente que el procedimiento analítico es adecuado al propósito. La validación y/o verificación de los métodos, para una determinada aplicación y en las condiciones particulares del laboratorio, permite un conocimiento de las características de funcionamiento del método, proporciona un alto grado de confianza y seguridad en el mismo y en los resultados que se obtienen, e incorpora lo necesario para establecer la trazabilidad metrológica de los resultados.

La validación de un método analítico es un proceso riguroso, dependiente de la técnica y de otros factores tales como la matriz a analizar, analitos, rangos de concentración, parámetros a evaluar, infraestructura disponible, etc. Un proceso de validación y/o verificación de un método debiera incluir las siguientes actividades:

1. Recopilación de información: estado del arte del método analítico a validar, normativas nacionales e internacionales
2. Insumos: búsqueda, selección, cotización y adquisición
3. Diseño esquema de validación: Este diseño permite determinar las variables y condiciones de las etapas prácticas del trabajo a desarrollar y considera:
   - Definición de los requerimientos del método
   - Pretratamiento de las muestras
   - Puntos críticos del método
   - Condiciones para evaluar precisión
   - Condiciones para evaluar exactitud
   - Uso de muestras controles (MRC, spike, etc)
   - Números de experimentos/muestras/analistas a realizar
4. Proceso de Validación y/o verificación: corresponde a la etapa experimental donde se evalúan, optimizan y definen los parámetros relevantes que permitirán dar cumplimiento con los requerimientos técnicos establecidos.

ALCANCE

Esta guía detalla el plan de validación de métodos de análisis químico para productos hidrobiológicos, siguiendo las directrices de las principales normativas internacionales vigentes. Debido a la importancia comercial y estratégica, especial atención se ha puesto en los requerimientos de la guía de validación de la Comunidad Europea CE/657/2002 Eurachem. Sin embargo, también se ha consultado las directrices de guías específicas (ISO) para la estimación de algunos de los parámetros de validación.

Tal como se discute en secciones posteriores de esta guía, Sernapesca comprende que existen varias aproximaciones a la validación de métodos analíticos las cuales, si disponen de suficiente base estadística, química y metrológica; podrían complementar y/o reemplazar algunas de las directrices de esta guía. El objetivo de este documento es sentar las bases mínimas de un proceso de validación de métodos analíticos confirmatorios para productos hidrobiológicos. Por lo tanto, quedan fuera del alcance de esta guía los métodos de *screening*.
La validación de los métodos analíticos debe tener un alcance finito y acotado, el cual debe quedar explícito en el informe de validación. Este alcance debe considerar los tanto los tipos de productos (p.ej: choritos, crustáceos, etc.) como su presentación (p.ej: freco, congelado, salado, etc.)
En el presente documento se discuten y definen los siguientes lineamientos del proceso de validación:
Entre paréntesis se indican las guías que fueron consultadas para adoptar métodos estadísticos y/o criterios de validación

4. Sesgo/Recuperación (ISO 5725, CE/657/2002, VIM)
5. Precisión Repetibilidad & y Reproducibilidad (ISO 5725, ISO 11352, ISO 7870)
6. Linealidad (ISO 8466, ISO 11095)
7. Límite de detección y Límite de cuantificación (IUPAC)
8. Incertidumbre (ISO 98:2008)
9. Robustez (CE/657/2002)

Se describe el diseño experimental mínimo para evaluar cada parámetro analítico. Este diseño está basado en principios estadísticos formales e intenta ajustarse, en forma general, a los procedimientos de trabajo habitual y de rutina del laboratorio.

Métodos no rutinarios o especiales deben evaluarse caso a caso y presentar la evidencia documental a Sernapesca para su evaluación.

Se entregan las directrices generales de los métodos estadísticos que son utilizados en la validación. Si bien estos métodos se ajustan a la normativa internacionalmente aceptada (Guías ISO, Eurachem, etc.), éstos deben ser aplicados e interpretados considerando el criterio químico como eje principal de discusión. Si bien tanto el diseño experimental como los métodos estadísticos propuestos están basados en guías internacionales, el laboratorio podrá utilizar métodos estadísticos no contemplados en este procedimiento, siempre y cuando estén debidamente documentados y que cumplan con los requisitos de aceptabilidad y pertinencia. Aunque esta guía entrega detalles de los cálculos de la mayoría de los métodos estadísticos propuestos, se asume que el(la) responsable de planificar y evaluar la validación posee un conocimiento básico de estadística aplicada en química analítica. En concreto, debe estar familiarizado(a) con la aplicación e interpretación de los siguientes conceptos:

- Medidas de tendencia central: promedio, moda y mediana
- Medidas de dispersión: desviación estándar, coeficiente de variación, varianza.
- Regresión lineal aplicada a calibración química.
- Análisis de varianza (ANOVA): es necesario conocer cómo implementarlo en un software estadístico o planilla de cálculo
- Distribución Normal
- Pruebas de hipótesis: conocer el concepto de un test estadístico y su interpretación.

No es necesario conocer los detalles teóricos de cada técnica, sino más bien estar familiarizado con los conceptos. Esta guía entrega directrices de cómo interpretar, con criterio químico, los resultados obtenidos al aplicar cada uno de los métodos. La bibliografía adjunta posee abundantes referencias a estadística básica aplicada en química analítica.

1. LINEALIDAD DE LA CURVA DE CALIBRACIÓN

Los objetivos del estudio de linealidad son i) establecer el rango de concentraciones de trabajo y ii) demostrar que, en el rango establecido, la respuesta instrumental (Y) es proporcional a la concentración del analito(X), es

decir, es posible modelar Y en función de X a través de un modelo lineal simple de la forma:

$$y = \beta_0 + \beta_1 x + \epsilon$$

donde β_0 es un intercepto, β_1 es la pendiente y ϵ es un error aleatorio que se asume independiente entre cada nivel de calibración, y puede modelarse mediante una distribución Normal con media $\mu = 0$ y varianza constante σ^2, es decir, datos *homocedásticos*.

Desde el punto de vista puramente estadístico, no se puede demostrar que el modelo lineal es el *verdadero* modelo de calibración, sino que solamente podremos concluir que, para los datos obtenidos experimentalmente, el modelo lineal es *adecuado* o *razonable* para los fines u objetivos de la calibración. Los modelos matemáticos o estadísticos son simplemente una abstracción de la realidad, no es una descripción absoluta. Por lo tanto, no existe ninguna prueba estadística que permita afirmar que el modelo de calibración es lineal.
Existen varias aproximaciones para evaluar que la linealidad de una curva de calibración es adecuada o razonable. Esta guía propone el uso de dos test estadísticos formales de linealidad:

10. Test de Carencia de Ajuste (*Lack of Fit*) descrito en la guía ISO 11095
11. Test de Mandel descrito en la guía ISO 8466-1.

Antes de describir estos tests, debemos hacer la siguiente acotación: En estadística un modelo lineal es aquel en que sus parámetros son lineales. Por ejemplo, la curva de calibración lineal clásica $y = \beta_0 + \beta_1 x$ tanto β_0 y β_1 son lineales. Sin embargo, en un modelo exponencial $y = \gamma_0 e^{\gamma_1 x}$ el coeficiente γ_1 no lo es. Dicho esto, el modelo cuadrático de calibración $y = \beta_0 + \beta_1 x + \beta_2 x^2$ es lineal desde el punto de vista estrictamente estadístico. Sin embargo, debido al arraigo del concepto de linealidad en química analítica no modificaremos su interpretación.

Los diseños experimentales para aplicar ambos tests, deben ser evaluados en condiciones de repetibilidad.
Los estudios de linealidad deben tener en cuenta los requerimientos de los distintos mercados de destino de los productos hidrobiológicos. Por lo tanto, deberá evaluar si el rango de trabajo validado cubre todos los requerimientos. Si no fuese así, el laboratorio deberá establecer y validar

distintos rangos de trabajo lo que podría incluir, salvo que se aplique una operación de dilución de la muestra, llevar a cabo las pruebas de linealidad en cada uno de los rangos.

Nota sobre el coeficiente de correlación como prueba de linealidad

La evidencia teórica y empírica es abrumadora en contra del uso del coeficiente de correlación como una prueba formal de linealidad, tal como puede apreciar las siguientes referencias [REF]. En algunos métodos analíticos esta propiedad tiene un sustento teórico. Por ejemplo, bajo ciertas condiciones existe una relación lineal entre absorbancia y concentración (Ley de Lambert-Beer) o entre potencial y el logaritmo de la concentración (Ecuación de Nernst), por lo que es razonable que en un cierto rango de concentración el modelo lineal sea adecuado. Es cierto, que existen desviaciones en los extremos, por lo tanto, el objetivo es establecer el rango de concentraciones donde esta propiedad se cumple.

Por lo tanto, no tiene sentido tratar de obtener el mayor número de "9s" del coeficiente de correlación. La linealidad no es una propiedad cualitativa, una curva con r = 0,999 no es más lineal que otra con r = 0,99. Observe la siguiente figura, ambas curvas fueron simuladas (en lenguaje R) con el mismo intercepto y pendiente siguiendo estrictamente el modelo lineal $y = \beta_0 + \beta_1 x$, sin embargo, tienen coeficientes de correlación distintos. La curva de la izquierda no es más lineal que la de la derecha ya que ambas fueron simuladas a partir del mismo modelo, la única diferencia es que la de la derecha tiene una mayor dispersión de los puntos alrededor de la recta, pero no por ello es "menos lineal". De hecho, ambas lo son, pues fueron simuladas con los mismos parámetros (excepto σ).

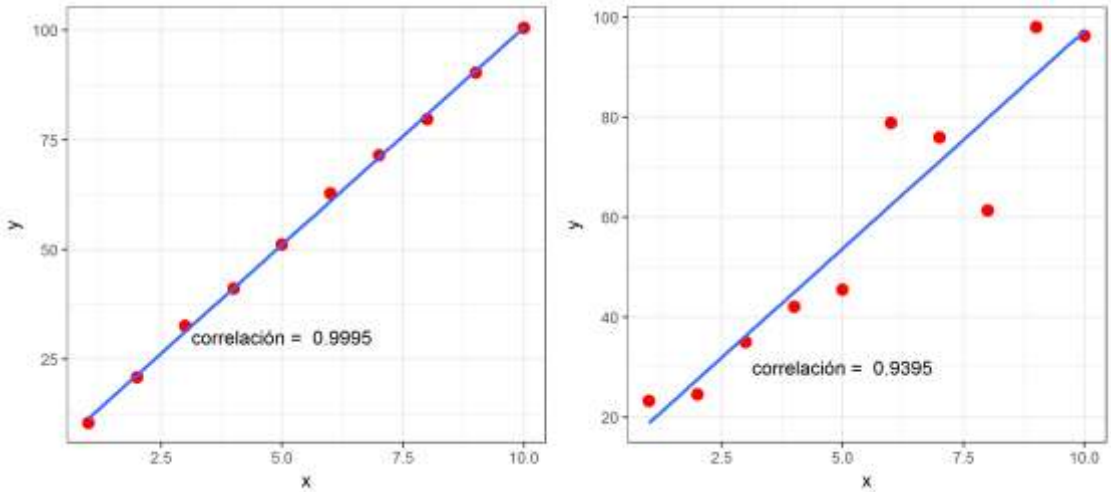

Figura 1. Simulación de curvas perfectamente lineales

Otro problema con el coeficiente de correlación como prueba de linealidad, radica en que un mal diseño de la curva de calibración podría, eventualmente, "inflar" artificialmente este parámetro. Es posible obtener coeficientes de correlación "altos" inclusive con datos que, a simpe vista, no revelan una relación lineal entre X e Y. Más grave aún, es posible obtener un alto coeficiente de correlación donde no existe absolutamente ninguna correlación entre X e Y, mediante un mal de diseño de la curva de calibración. Por ejemplo, en la siguiente figura se muestra a la izquierda que la correlación entre x e y es prácticamente 0. Sin embargo, cuando incluimos un punto *influyente*, muy alejado de la nube de puntos a baja concentración, el coeficiente de correlación r = 1 (derecha).

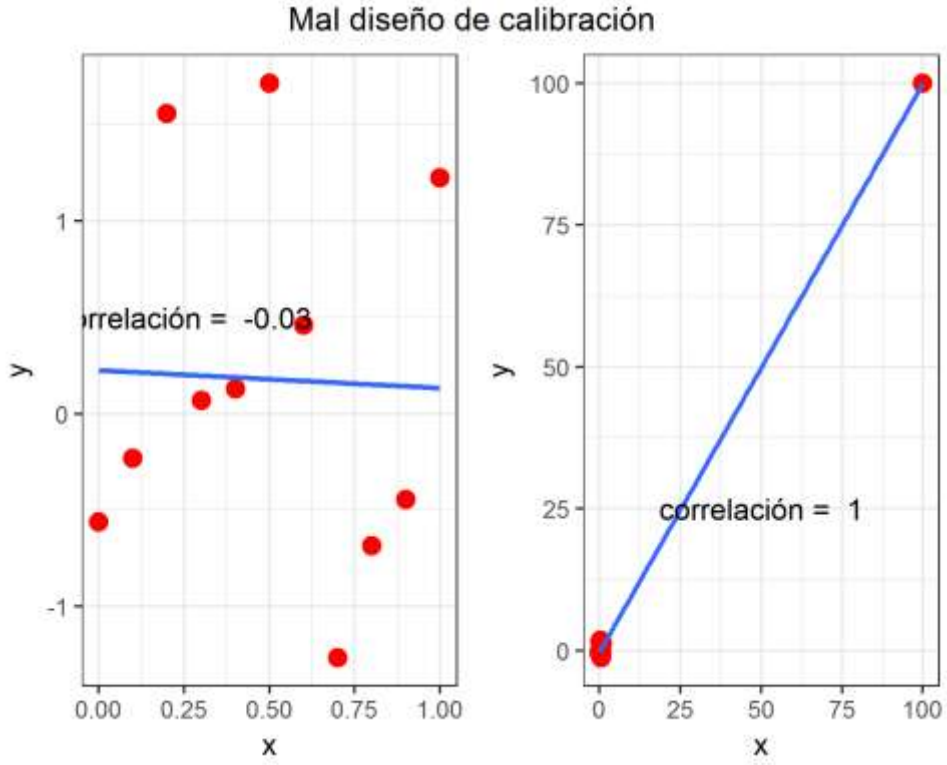

Figura 2. Punto influyente en la calibración

El rol que juega el coeficiente de correlación en calibraciones químicas está relacionado con la incertidumbre de calibración. Ellison demuestra que a mayor r, menor es la incertidumbre de calibración [17].

Nota sobre curvas de calibración que pasan por el origen

Algunos softwares de equipos de medición (AAS, cromatógrafos, etc.) permiten ajustar varios modelos de calibración lineales y no lineales. De especial interés es el modelo conocido como *lineal-origen* el cual no incluye un intercepto:

$$y = \beta_1 x + \epsilon$$

Es decir, fuerza a la curva de calibración a pasar por el origen (0,0). Este hecho tiene algunas consideraciones metrológicas que vale la pena mencionar:

- Es posible que un blanco analítico del cual se asume que la concentración del analito es < LoD, presente una señal instrumental significativa, especialmente en métodos AAS o ICP.
- En teoría, en los métodos cromatográficos una muestra blanco no debiera presentar ningún peak en el tiempo de retención del analito, es decir, sólo se debiera observar línea base o ruido instrumental.
- Incluir el punto (0,0) en una curva de calibración sin realmente llevar a cabo una medición no es correcto desde el punto de vista metrológico. Además, ya que los extremos de la curva de calibración tienen una gran influencia en la estimación de la pendiente e intercepto, incluir el punto (0,0) podría generar sesgos en la predicción de la concentración de muestras de baja concentración.

En cualquier caso, se recomienda, en primer lugar, evaluar si la curva pasa por el origen a través de una prueba estadística formal del intercepto:

$$\begin{aligned} H_0\colon\ \beta_0 &= 0 \\ H_1\colon\ \beta_0 &\neq 0 \end{aligned}$$

Esta prueba es muy sencilla y está incorporada en todos los softwares estadísticos y planillas de cálculo. A continuación, se muestra el análisis estadístico en MS Excel de una curva de calibración que sí incluye el intercepto:

	Coeficientes	*Error típico*	*Estadístico t*	*p-value*	*Inferior 95%*	*Superior 95%*
Intercepción	10,0	8,4	1,2	0,3	-11,6	31,5
X	10,1	0,2	43,3	1,E-07	9,5	10,7

Figura 3. Prueba formal del intercepto

Para interpretar la prueba estadística sobre el intercepto, es necesario evaluar el *p-value* correspondiente al término *intercepción*. En este ejemplo, el *p-value* = 0.3, el cual es mayor que el nivel de significancia 0.05, por lo

tanto, no hay evidencias en contra de la hipótesis nula H_0: $\beta_0 = 0$ y es razonable asumir que la curva pasa por el origen.
Una vez que se ha demostrado que esta hipótesis es plausible, entonces, es posible ajustar un modelo de calibración *lineal-origen*, es decir, sin intercepto.
Por lo tanto, la recomendación de esta guía es que el primer punto de calibración corresponda al blanco analítico incluyendo su señal instrumental cuando corresponda.
Test de Carencia de Ajuste. Guía ISO 11095

Para llevar a cabo esta prueba estadística es necesario preparar replicados independientes (genuinos) para cada nivel de concentración. Este test establece dos hipótesis a contrastar:

- H_0: No hay carencia de ajuste, el modelo lineal es adecuado/razonable para explicar la relación entre la respuesta instrumental (Y) y Concentración (X).
- H_1: Existe carencia de ajuste, el modelo lineal no es adecuado para explicar la relación entre la respuesta instrumental (Y) y Concentración (X).

Los detalles estadísticos del test de carencia de ajuste pueden consultarse en la bibliografía adjunta. Brevemente, en este test la suma de cuadrados de los residuos $SS_r = \sum_{i=1}^{n} e_i^2$ es particionada en dos componentes:

$$SS_r = SS_{\text{error puro}} + SS_{\text{lack-of-fit}}$$

donde $SS_{\text{error puro}}$ representa la variabilidad debido al error puramente experimental (de ahí la necesidad de tener replicados genuinos de cada calibrante, es un "análogo" a la repetibilidad intra-calibrante) y $SS_{\text{lack-of-fit}}$ representa la falta de ajuste del modelo elegido, en este caso del modelo lineal.

Cuando estas sumas de cuadrados se dividen por sus respectivos grados de libertad, se obtienen las varianzas de error puro σ_p^2 y de carencia de ajuste (lack-of-fit) σ_{lof}^2, respectivamente.
Si el modelo lineal es adecuado, entonces, estas varianzas deben ser iguales. Para probar esta hipótesis de igualdad de varianzas se utiliza un test F, donde:

$$F_{calculado} = \frac{\sigma_{lof}^2}{\sigma_p^2}$$

el cual se compara con un F de tabla con $n-2$ grados de libertad en el numerador y $n(k-1)$ en el denominador. k es el número de replicados independientes en cada uno de los n niveles de calibración.

Si bien este test puede ser implementado fácilmente en una hoja de cálculo, la mayoría de los softwares estadísticos lo incorporan en forma rutinaria. Al igual que el test de Mandel, los softwares estadísticos informan el *p-value* de la carencia de ajuste, cuya interpretación es la siguiente:

- Si *p-value* de la carencia de ajuste (*lack of fit*) > 0.05, entonces, no ha evidencias en contra de la hipótesis nula, por lo tanto, se concluye que el modelo lineal es razonable o adecuado.

- Si *p-value* de la carencia de ajuste (*lack of fit*) < 0.05, entonces, se rechaza la hipótesis nula, por lo tanto, se concluye que el modelo lineal NO es adecuado.

A continuación, se muestra un ejemplo de aplicación de este test. La siguiente tabla y figura muestran los datos de calibración de cloranfenicol en matriz leche obtenida por GC/MS-NCI (*Gas Chromatography/Mass Spectrometry - Negative Chemical Ionization*). Note que cada nivel de calibración $n=5$ está preparado en triplicado ($k=3$) totalizando 15 calibrantes independientes.

Tabla 1. Datos de calibración test de carencia de ajuste

Replicado	x.caf	y.caf
1	0	88
2	0	154
3	0	512
1	0	7714
2	0	7726
3	0	8043
1	0	15292
2	0	14947
3	0	15063
1	1	22611
2	1	22945
3	1	22772
1	1	30280
2	1	30222
3	1	30089

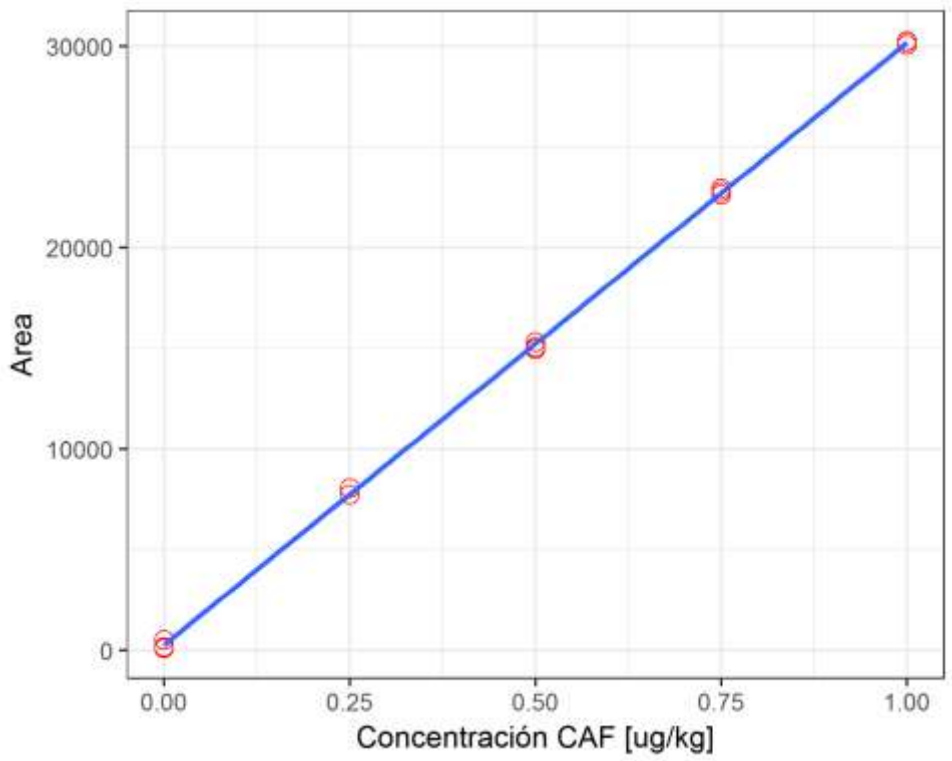

Figura 4. Curva de calibración CAF. Test de carencia de ajuste

La tabla siguiente muestra el análisis estadístico de esta calibración realizado en el software *R*.

Pendiente e intercepto de calibración

term	estimate	std.error	statistic	p.value
(Intercept)	262	78	3	0.005
x.caf	29936	127	236	0.000

A continuación, se muestra el test de carencia de ajuste: Tabla 2. Tabla de test Carencia de Ajuste

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x.caf	1	1680303154	1680303154	54131	0.00
Residuals	13	392486	30191		
Lack of fit	3	82074	27358	1	0.48
Pure Error	10	310412	31041		

Para interpretar el test de carencia de ajuste hay que evaluar el *p-value* correspondiente al "*Lack of fit*", el cual aparece bajo la columna *Pr(>F)* = 0.48.

La interpretación tradicional de esta prueba estadística concluiría que dado que el *p-value* > 0.05, entonces, no hay evidencias en contra de la hipótesis nula. El modelo lineal es adecuado para modelar los datos de calibración. Una discusión más profunda de esta conclusión se presenta a continuación, de la descripción del test de Mandel.

Test de Mandel ISO 8466-1

Esta prueba estadística es bastante sencilla y está basada en la comparación entre el modelo de calibración lineal y un modelo alternativo. Por lo tanto, no es una prueba absoluta, sino relativa a la elección del modelo alternativo. Requiere al menos n = 6 calibrantes (sin replicado).

En general, el test de Mandel utiliza el modelo de calibración cuadrático para compararlo con el modelo lineal:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Los detalles estadísticos pueden consultarse en la bibliografía. Describiremos brevemente el procedimiento, el cual puede ser implementado en una hoja de cálculo:

12. Primero calcule la suma de cuadrados de los residuos SS_r para cada uno de los modelos de acuerdo a la siguiente expresión:

$$SS_r = \sum_{i=1}^{n} e_i^2$$

Donde el residuo $e = y - \hat{y}$. y es la respuesta instrumental observada o experimental (áreas, absorbancias, etc.); $\hat{y}$, es la respuesta instrumental que predice el modelo (lineal o cuadrático) en cada una de las concentraciones de los calibrantes. Si observa la ecuación, el concepto de residuo es el mismo para cualquier modelo de calibración, es decir, ¿cuánto difiere lo que se observa experimentalmente con lo que predice el modelo?

Un buen modelo tiene residuos pequeños. Un residuo grande para cierto de nivel de concentración implica que existe una gran diferencia entre lo observado y lo que predice el modelo, por lo tanto, nos da pistas para detectar posibles valores anómalos o *outliers*.

2. Calcule la diferencia entre ambas sumas de cuadrado de los residuos, la del modelo no lineal SS_r^{no-lin} y la correspondiente al modelo lineal SS_r^{lin}:

$$D = SS_r^{no-lin} - SS_r^{lin}$$

3. Calcule el estadístico F calculado:

$$F = \frac{D}{SS_r^{no-lin}/(n-3)}$$

5. Obtenga el F de tabla para 1 grado de libertad en el numerador y $n-3$ para el denominador

6. Compare el $F_{calculado}$ con el F_{tabla} y decida en base a la siguiente regla:

- Si $F_{calculado} < F_{tabla}$ se concluye que no hay evidencias en contra de la hipótesis nula de linealidad del modelo. El modelo lineal es razonable o adecuado.

- Si $F_{calculado} > F_{tabla}$ se rechaza la hipótesis nula de linealidad del modelo. Los datos no son consistentes con la hipótesis de linealidad.

- Actualmente, todos los softwares estadísticos informan el *p-value* del test de Mandel, el cual contiene la misma información que la comparación del $F_{calculado}$ con F_{tabla}. Si este *p-value* < 0,05, entonces, se rechaza la hipótesis nula de linealidad. Por lo tanto, este modelo no sería adecuado para describir los datos de calibración.

A continuación, se muestra un ejemplo de aplicación del test de Mandel analizado a través de software estadístico.

La tabla y figura siguientes muestran los datos de calibración:

Curva de Calibración

x	y
0	-0.007
10	0.071
20	0.146
30	0.212
40	0.274
50	0.334
60	0.385

70	0.430
80	0.473
90	0.511
100	0.546

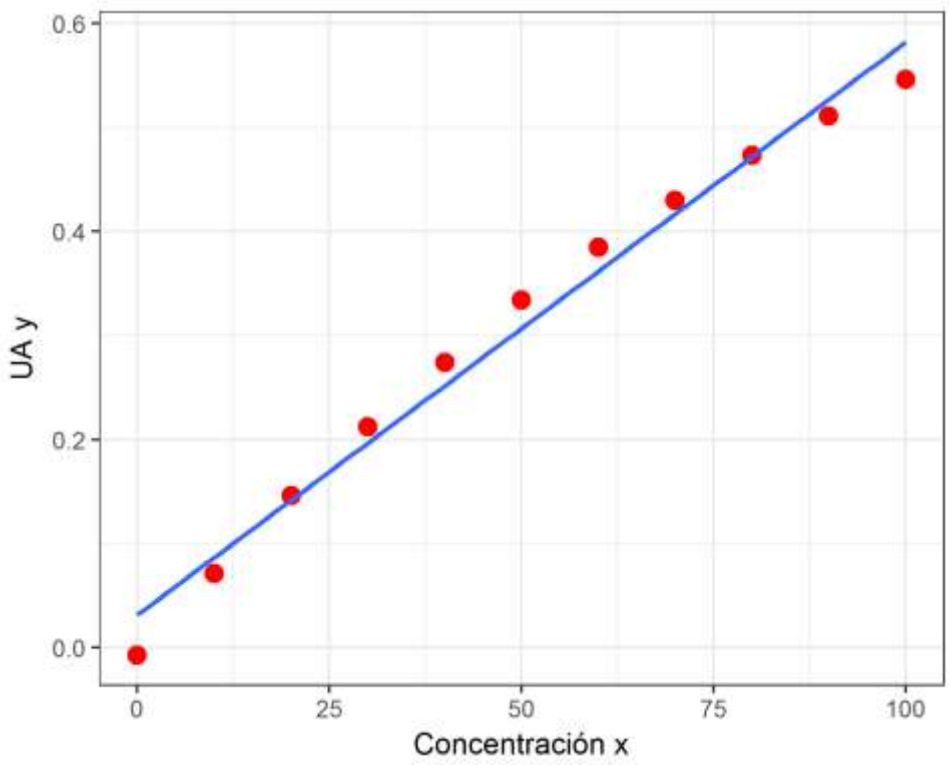

Figura 5. Curva de calibración

A simple vista se observa la no linealidad de la curva de calibración. Veamos que nos indica el Test de Mandel obtenido en un software estadístico:

```
Analysis of Variance Table

Model 1: y ~ x
Model 2: y ~ x + I(x^2)
  Res.Df      RSS Df Sum of Sq      F    Pr(>F)
1      9 0.0054816
2      8 0.0000186  1  0.005463 2351.1 3.621e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En la tabla anterior el valor $F = 2351$ corresponde al $F_{calculado}$ el cual es comparado internamente por el software con el F_{tabla} entregando, finalmente, el *p-value* $Pr(> F) = 3.62e - 11$. Ya que este *p-value* << 0.05, la evidencia en contra de la hipótesis nula de linealidad es abrumadora.

Si utiliza una planilla de cálculo para implementar el test de Mandel (en Excel), la forma más sencilla es ajustar una regresión cuadrática mediante el menú *Herramientas para análisis*. Una vez obtenida la tabla de resultados, el *p-value* del test de Mandel es aquel que acompaña al término cuadrático del modelo.

Algunas consideraciones para interpretar los tests de linealidad

13. Si el *p-value* > 0.05, esto no significa que el modelo de calibración sea lineal. Sólo indica que, con los datos experimentales, no existe evidencia en contra de la hipótesis de linealidad. Bajo el paradigma de las pruebas de hipótesis, las hipótesis nulas no se aceptan. Lo único que se puede extraer como conclusión es que el modelo lineal es adecuado, es razonable para modelar los datos de calibración.

14. Existen muchos modelos de calibración que podrían ser idóneos, estos tests nos indican si el que hemos elegido para modelar los datos es razonable/adecuado, sin embargo, no nos indica que sea "EL" modelo perfecto.

15. ¿Qué tiene de especial el valor de corte 0,05? Absolutamente *nada*, es completamente arbitrario ¿Cuál sería la conclusión si el *p-value* fuese 0,04999 ó 0,05001? Sería un test totalmente inconcluyente.

16. Lamentablemente, esta es una dicotomía que desde hace mucho tiempo ha sido objeto de varias críticas. Se sugiere la lectura de las siguientes referencias sobre la interpretación y controversia de los *p-values* en investigaciones científicas:

- Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA's Statement on p-Values: Context, Process, and Purpose *The American Statistician Volume 70, 2016 - Issue 2* link

- M.Baker Statisticians issue warning over misuse of P values *Nature 531, 151 (10 March 2016)* link

- Singh Chawla D. Big names in statistics want to shake up much-maligned P value. *Nature. 2017 Jul 26;548(7665):16-17* link

En conclusión, no se puede demostrar que el modelo de calibración es lineal. Sino que sólo es posible concluir que es razonable o adecuado, tal como cualquier tipo de modelo cuantitativo propuesto en ciencias. Es sólo eso: *un modelo*.

5. En general, para la evaluación de linealidad se debe establecer al menos $n = 6$ niveles de calibración en forma equiespaciada (por ejemplo: 10-20-30-40-50-60 ppb). Cada uno de los calibrantes puede ser sometidos a replicados de lectura instrumental, sin embargo, esto no significa un aumento en los puntos de la curva de calibración. Por ejemplo, $n = 6$ calibrantes leídos en duplicado no implica que el número total de puntos de calibración sea igual a 12. Sí es correcto promediar los duplicados de lectura para cada nivel de concentración, por lo tanto, el número de calibrantes sigue siendo $n = 6$.

Método de calibración con adición de estándar interno

En general, especialmente en los métodos cromatográficos, es recomendable el uso de un estándar interno el cual debe ser agregado tanto a los calibrantes como a las muestras, en una cantidad fija y conocida. Este compuesto, debe tener las siguientes características:

17. Debe tener una estructura química similar mensurando de interés. Por ejemplo, en la determinación de cloranfenicol por cromatografía puede utilizarse florfenicol como estándar interno.

18. No debe estar presenta en la muestra

19. Idealmente debe tener una respuesta adecuada para el tipo de detector

20. La señal analítica (tiempo de retención, m/z, etc.) del IS debe estar completamente resuelta. En el caso de MS, si hubiese interferencias de la matriz en la señal m/z del IS, éstas deben ser evaluadas y corregidas.

Idealmente el estándar interno de adicionarse desde el inicio del proceso analítico, con el fin de corregir las pérdidas del analito en todas las etapas del método. Si es adicionado justo antes de la inyección en el equipo, sólo corregirá errores durante este proceso.

La curva de calibración con estándar interno se construye a partir de las razones de señales analíticas (Y) y las razones de las concentraciones (X), entre el analito y el IS tal como lo muestra la siguiente expresión:

$$\underbrace{\frac{R_{\text{analito}}}{R_{IS}}}_{y} = \beta_0 + \beta_1 \underbrace{\frac{C_{\text{analito}}}{C_{IS}}}_{x}$$

donde R esla respuesta instrumental. Ya que a la muestra problema también se le adicionó la misma cantidad de IS, entonces es posible despejar de la ecuación anterior la concentración de la muestra a partir de la siguiente expresión:

$$C_{muestra} = \left(\frac{R_{\text{muestra}}}{R_{IS}} - \beta_0\right)\frac{C_{IS}}{\beta_1}$$

Si es utilizada en forma correcta, la calibración con estándar interno mejorará la precisión y minimizará el sesgo de la medición si se compara con la calibración tradicional, especialmente en aquellos métodos analíticos con varios pasos de extracción y separación.

Nota para métodos cromatográficos acoplados a espectrometría de masas

Como una sugerencia, mas no un requerimiento, en métodos cromatográficos acoplados a espectrometría de masas se aconseja el uso de compuestos marcados isotópicamente como estándares internos. De esta manera se asegura una correcta corrección de pérdidas del analito durante el proceso analítico (extracción, separación, derivatización, etc.) obteniendo valores de concentración con bajo sesgo y menor incertidumbre.

Por ejemplo, para la determinación de cloranfenicol, se encuentra disponible comercialmente el compuesto hexadeuterado D6-Cloranfenicol. Es muy importante definir los iones que serán utilizados para la cuantificación ya que algunas razones m/z podrían solaparse.

Nota para métodos ICP-OES e ICP-MS.

En el caso de la determinación de metales por ICP también se recomienda, mas no es un requerimiento, utilizar un estándar interno adecuado al analito. En el caso de ICP-MS es posible utilizar un metal con una razón m/z cercana al metal de interés, así como también, aunque a un costo muy elevado, utilizar una solución isotópicamente enriquecida. Es decir, para la determinación de Cd puede usarse, por ejemplo, alguno de sus isótopos (^{116}Cd) como IS en ICP-MS.

Nota sobre calibración en matriz (*matrix-matched calibration*)

Está documentado en métodos cromatográficos para la determinación de compuestos orgánicos la, eventual, presencia del efecto matriz durante todo el proceso analítico: extracción, *clean-up*, inyección (efecto protector de la matriz en GC), ionización en MS, etc. Se han desarrollado diversas estrategias para minimizar estos efectos: uso de estándares internos isotópicamente marcados, dispositivos especiales de introducción de muestras, calibración en matriz, etc. Si los estudios de validación sugieren la presencia de un efecto matriz significativo cuando se utiliza calibración estándar (en solvente), entonces, esta guía sugiere el uso de calibración en matriz para minimizar estos efectos. Esta sugerencia aplica sólo a métodos cromatográficos.

2. PRECISIÓN

La literatura y las normas internacionales describen varios procedimientos para obtener la precisión de un método en condiciones de repetibilidad y reproducibilidad (r & R). El siguiente protocolo sigue las directrices de la Guía ISO 5725 parte 3 y permite obtener los parámetros de repetibilidad y precisión intermedia en un mismo estudio. En rigor, la reproducibilidad sólo puede estimada a través de un ensayo colaborativo entre laboratorios. También se propone la estimación de la precisión de repetibilidad a través del análisis de muestras en duplicado, la cuales pueden corresponder a datos históricos de QAQC.

Esta guía también describe cómo estimar el límite de repetibilidad (Reproducibilidad), el cual puede ser utilizado como criterio de aceptación para mediciones en duplicado.
Estimación de la precisión de acuerdo a la Guía ISO 5725 parte 3

El protocolo propuesto por esta guía está basado en la estimación de la precisión mediante un diseño estadístico experimental en el cual es posible estudiar un número de *factores* que, eventualmente, podrían influir en la precisión de un método analítico. Estos factores podrían corresponder, mas no circunscribirse a:

21. Analista
22. Instrumento
23. Días (tiempo)
24. etc.

De acuerdo a este diseño experimental es posible obtener, bajo un mismo estudio, los parámetros de repetibilidad y precisión-intermedia. La desventaja de esta propuesta subyace en que es imposible conocer todos los factores que podrían influenciar la precisión del método, además, que al aumentar el número de factores, el tamaño del estudio (número de experimentos) crece rápidamente.

Desde el punto de vista estadístico, el modelo utilizado para la estimación de la precisión corresponde a un modelo lineal de efectos aleatorios, cuyos parámetros de precisión son obtenidos a partir de un análisis de varianza (ANOVA).
El protocolo básico consiste en llevar a cabo análisis de replicados de una única muestra bajo las distintas condiciones experimentales que se

describen en la matriz de diseño experimental. Idealmente, esta única muestra debe corresponder a una muestra incurrida. Si no es posible obtenerla, pueden preparase mediante fortificación de una muestra blanco la cual debe ser sometida a una exhaustiva homogenización. No es necesario disponer de materiales de referencia para estudios de precisión.

El siguiente protocolo detalla el estudio de un único factor. Para diseños que consideren 2 o más factores, puede consultarse la guía ISO 5725 parte 3:

25. Definir el factor y los niveles a estudiar. Por ejemplo, se estimará la precisión del método durante cinco días no consecutivos. En este caso el factor de precisión intermedia es *Tiempo* y los niveles del factor son 5 (5 días no consecutivos). Una vez definido el factor, todos los demás factores no estudiados se deben mantener constantes (mismo analista, mismo equipo, etc).

26. En cada uno de los niveles del factor se analizarán n replicados de una única muestra homogenizada. No es necesario conocer la concentración de la muestra a utilizar, sin embargo, sería factible combinar los estudios de precisión con la evaluación de recuperación si se conociera la concentración de la muestra fortificada. A continuación, se muestra un ejemplo de un diseño experimental que considera el estudio del factor Tiempo, con 5 niveles (5 días no consecutivos) y $n = 6$ replicados independientes. Todos los análisis deben ser llevados a cabo por un mismo analista en un mismo equipo:

Tabla 3. Diseño experimental de precisión 1 factor

Día 1	Día 2	Día 3	Día 4	Día 5
Replicado 1	Replicado 1	Replicado 1	Replicado 1	Replicado 1
Replicado 2	Replicado 2	Replicado 2	Replicado 2	Replicado 2
Replicado 3	Replicado 3	Replicado 3	Replicado 3	Replicado 3
Replicado 4	Replicado 4	Replicado 4	Replicado 4	Replicado 4
Replicado 5	Replicado 5	Replicado 5	Replicado 5	Replicado 5
Replicado 6	Replicado 6	Replicado 6	Replicado 6	Replicado 6

4. Una vez obtenidos los datos experimentales, se debe realizar un análisis de varianza de un factor de acuerdo a las directrices de la guía ISO 5725-3. Todos los softwares estadísticos y planillas de cálculo incorporan esta herramienta. A continuación, se muestra, como ejemplo, datos experimentales del diseño propuesto (Factor Tiempo) y la tabla de análisis de varianza correspondiente:

Tabla 4. Diseño experimental de precisión 1 factor. Resultados experimentales en [ug/kg]

Replicado	Día 1	Día 2	Día 3	Día 4	Día 5
1	47.2	52.3	52.0	53.5	51.9
2	48.8	43.7	50.6	47.6	46.6
3	57.8	46.6	47.2	44.7	59.2
4	50.4	47.8	58.9	48.9	55.8
5	50.6	56.1	52.5	44.9	49.3
6	58.6	51.8	40.2	46.4	61.3

Antes de aplicar el análisis de varianza se sugiere llevara cabo un examen inicial de los datos con el fin de identificar patrones o valores anómalos que podrían distorsionar el análisis estadístico. A continuación, se muestra un gráfico de medias el cual indica que, aparentemente, hay consistencia entre las mediciones de los distintos días estudiados:

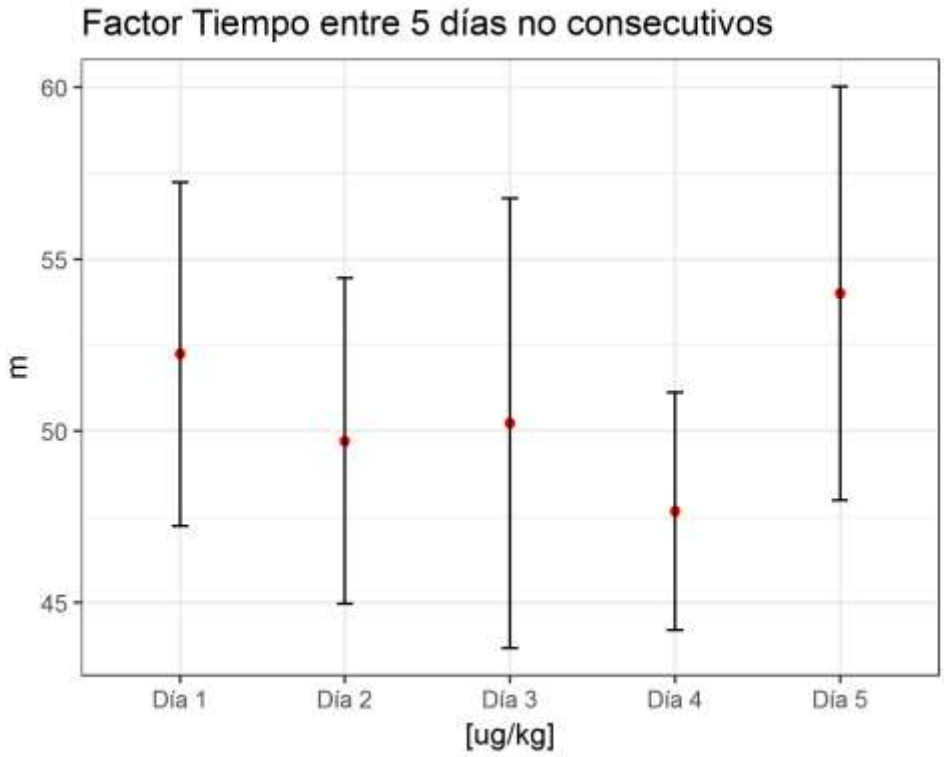

Figura 6. Boxplot de datos de precisión

En la eventualidad que los datos muestren valores atípicos(*outliers*) cuya calidad analítica no está en duda, sino que son un fiel reflejo de la

variabilidad real del método, éstos no deben eliminarse, pues se podría subestimar la precisión. En estos casos, para corroborar los cálculos de precisión se sugiere utilizar la técnica de análisis de varianza robusta, la cual está fuera del alcance de esta guía, pero puede consultarse en la bibliografía y descargar un *add-in* de Excel para llevar a cabo los cálculos en el siguiente link.

Nota: Salvo en algunos métodos no destructivos (p.e: Activación neutrónica para metales), es muy difícil separar la variabilidad analítica de repetibilidad (intra día, del mismo analista, etc.) de la heterogeneidad de la muestra. Por lo tanto, estos componentes siempre estarán incorporados en forma simultánea.

La siguiente tabla muestra el ANOVA del ejemplo de estudio de precisión:

Tabla 5. Tabla ANOVA estudio de precisión factor Tiempo

Factor de variabilidad	Suma de cuadrados	Grados de libertad	Cuadrados medios	F calculado	p value
Dia	4	142	35.5	1.41	0.26
Residuals	25	630	25.2		

Con los estadísticos de la tabla ANOVA, según se describe en la Guía ISO 5725 Parte 3, se obtienen los parámetros de precisión: repetibilidad y precisión intermedia:

27. Repetibilidad s_r se estima a partir de el cuadrado medio (CM) de los residuos (*residuals* o *within*): $s_r = \sqrt{CM_{Residuals}}$. En este ejemplo, se obtiene una repetibilidad de $s_r = 5\ \mu\text{g/kg}$.

28. Repetibilidad intermedia entre días $s_{entre-dias}$ se calcula a partir de las diferencias entre los cuadrados medios entre Días y los residuos (*Residuals*):

$$s_{entre-dias} = \sqrt{\frac{CM_{Dias} - CM_{Residuals}}{m}}$$

donde m es el número de replicado independientes realizados en cada día. En este ejemplo, se obtiene:

$$s_{entre-dias} = \sqrt{\frac{CM_{Dias} - CM_{Residuals}}{m}} = \sqrt{\frac{35.5 - 25.2}{6}} = 5.2\ \mu\text{g/kg}$$

5. Precisión intermedia s_I corresponde a la suma cuadrática entre los dos componentes arriba calculados: $s_I = \sqrt{s_r^2 + s_{entre-dias}^2}$. En este ejemplo se obtiene $s_I = 5.2\ \mu\text{g/kg}$.

La guía ISO 5725-3 es explícita en ejemplificar diversos diseños experimentales con más factores. Para estos diseños, se sugiere el uso de software estadístico que incorpore herramientas para modelos de efectos aleatorios, ya que es posible obtener directamente las desviaciones estándar de precisión de repetibilidad y intermedia. Puede encontrar estos métodos bajo el nombre de "Estudios r&R" (*repeatability & Reproducibility*). La gran ventaja de este tipo de estudios, es que permite estimar la contribución de cada uno de los factores a la precisión global (componente de varianza), por lo tanto, es una información valiosa que permite al laboratorio focalizar los esfuerzos en disminuir el efecto sólo de aquellos factores que más aportan a la variabilidad.

Nota: Los métodos propuestos por la guía ISO 5725-3, rigurosamente hablando, están dirigidos a la obtención de la reproducibilidad mediante estudios colaborativos. En esta guía sólo se aborda la estimación de la precisión intermedia como una estimación de la precisión intra-laboratorio.

Estimación de la precisión de repetibilidad $\boldsymbol{s_r}$ mediante análisis de duplicados de QAQC.

Es posible obtener la desviación estándar de repetibilidad a través del análisis estadístico de duplicados. Para aplicar este método es necesario disponer de al menos $n = 25$ datos correspondientes a muestras analizadas en duplicados independientes, es decir, dos porciones de muestra que han sido sometidas a todo el proceso analítico en condiciones de repetibilidad (mismo analista, mismo instrumento, el mismo día, etc.). Las muestras pueden corresponder a muestras de clientes, controles internos, materiales de referencia, rondas de intercomparación, etc. Lo importante es que los duplicados deben haber sido obtenidos bajo condiciones de repetibilidad.

La gran ventaja de esta metodología es que no requiere de un diseño experimental *ad-hoc* pues los datos pueden ser obtenidos del QAQC del laboratorio, reflejando fielmente las distintas condiciones de trabajo en un largo período de tiempo.

La principal desventaja, es que no permite estimar la contribución de cada uno de los eventuales factores involucrados en el análisis de rutina del laboratorio: analistas, instrumentos, etc.

El cálculo es sencillo, la desviación estándar de repetibilidad s_r es obtenida a partir de:

$$s_r = \sqrt{\frac{\sum_{i=1}^{n}(x_{i1} - x_{i2})^2}{2n}}$$

donde x_{i1} y x_{i2} son los duplicados de la muestra i y n es el número de muestras.

La siguiente tabla muestra un ejemplo de $n = 25$ muestras analizadas en duplicado:

Tabla 6. Datos de n = 25 muestras en duplicado [ug/kg]

Muestra	Duplicado 1	Duplicado 2	Diferencia
QAQC 101	44.7	44.2	0.6
QAQC 102	45.3	45.8	-0.5
QAQC 103	46.6	45.9	0.7
QAQC 104	46.3	45.7	0.6
QAQC 105	46.7	47.3	-0.6
QAQC 106	47.9	47.3	0.6
QAQC 107	47.7	47.4	0.4
QAQC 108	47.3	48.4	-1.1
QAQC 109	48.0	48.8	-0.8
QAQC 110	48.5	49.2	-0.6
QAQC 111	49.8	49.5	0.3
QAQC 112	49.8	49.9	-0.1
QAQC 113	50.2	50.0	0.2
QAQC 114	50.5	50.3	0.2
QAQC 115	50.6	50.6	-0.1
QAQC 116	52.1	50.9	1.2
QAQC 117	51.9	51.6	0.4
QAQC 118	51.1	51.5	-0.4
QAQC 119	52.9	53.6	-0.7
QAQC 120	52.7	53.5	-0.8
QAQC 121	52.8	52.8	0.0
QAQC 122	53.6	53.5	0.1
QAQC 123	53.7	53.9	-0.3
QAQC 124	54.2	55.0	-0.8

QAQC 125	54.7	55.0	-0.3

Se sugiere llevar a cabo un análisis exploratorio y evaluar al menos la presencia de valores *outliers* y la homocedasticidad, es decir, variabilidad constante en todo el rango de concentración estudiado. Para lo cual observe Duplicado 1 v/s Duplicado2 en un gráfico de dispersión (la línea roja punteada representa la pendiente y el intercepto teóricos $\beta_1 = 1$ y $\beta_0 = 0$).

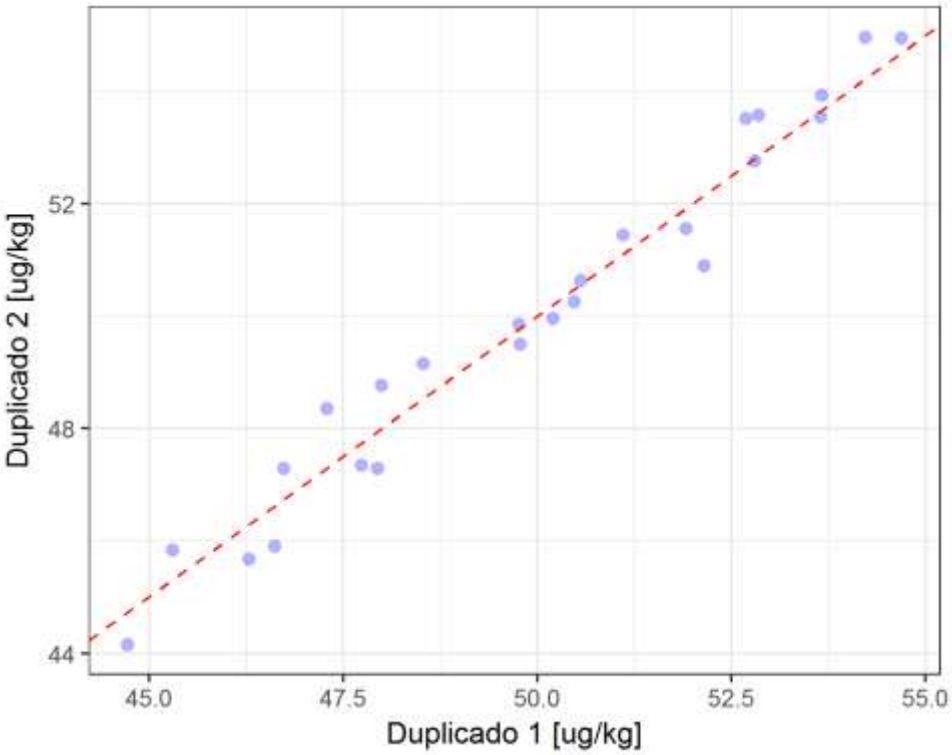

Figura 7. Gráficos de dispersión entre duplicados

No se advierte ningún patrón en los datos ni valores anómalos. Aunque está fuera del alcance de esta guía, como complemento al gráfico exploratorio, se sugiere llevar a cabo un análisis de residuos de la regresión de Duplicado 2 v/s Duplicado 1, el cual podría, eventualmente, revelar información adicional. Si este análisis evidenciara heterocedasticidad, es decir, a medida que aumenta la concentración existe mayor variabilidad (diferencia) entre duplicados, entonces, este método propuesto para la estimación de s_r debe ser modificado modelando la precisión de repetibilidad s_r en función de la concentración (véase el método de Howarth-Thompson), o bien, definir rangos de concentración en los cuales la variabilidad es constante. El siguiente gráfico muestra un ejemplo en que la dispersión entre duplicados es mayor a medida que aumenta la concentración:

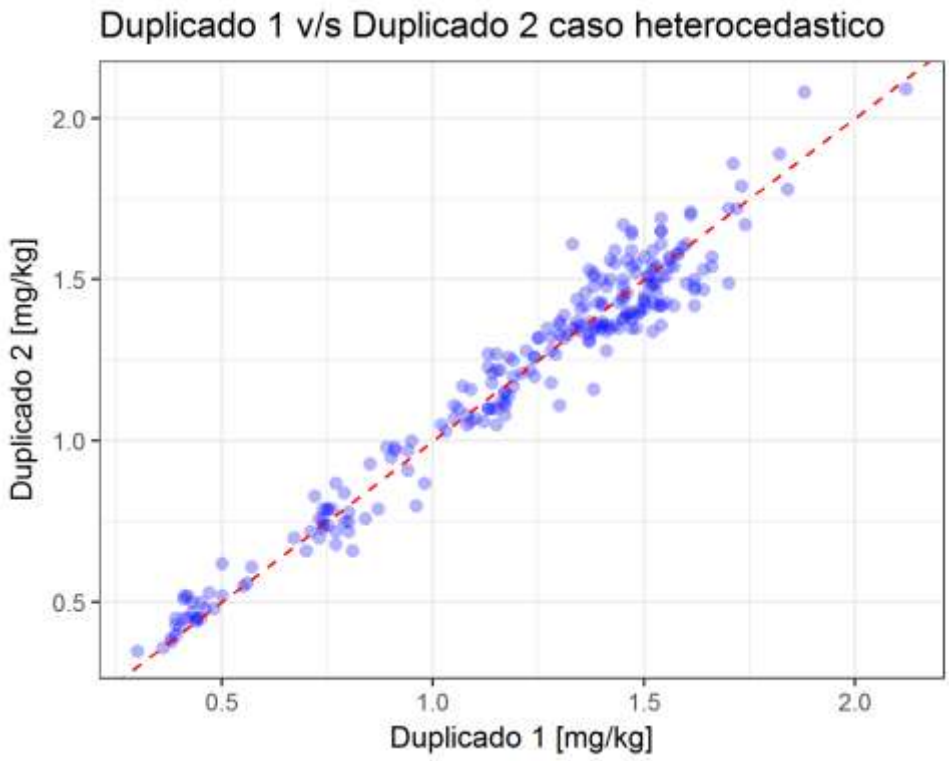

Figura 8. Gráfico de dispersión entre duplicados caso heterocedástico

Retomando el cálculo de la desviación estándar de repetibilidad para datos homocedásticos (Muestras QAQC) se obtiene $s_r = 0.41$ [mg/kg].
Estimación de la máxima diferencia tolerable entre duplicados

Los datos de precisión calculados pueden ser utilizados para establecer criterios de aceptación entre duplicados. La guía 5725 define el límite de repetibilidad r de la siguiente manera:

$$r = 2.8 \cdot s_r$$

y corresponde a la máxima diferencia, en valor absoluto, que puede tolerarse entre duplicados de análisis, obtenidos bajo condiciones de repetibilidad con un 95% de confianza, donde:

- s_r es la desviación estándar de repetibilidad, la cual da cuenta de la dispersión de las diferencias entre duplicados y puede ser obtenida a partir de uno de los métodos recién descritos u otro que esté bien documentado.
- El factor 2.8: este factor se relaciona con el grado de confianza con el cual se construirá este criterio de aceptación (95%). Nos indica que cuando se establece la

máxima diferencia tolerable, está permitido que el 5% de los duplicados estén fuera del límite r, y aún el sistema analítico se encontraría bajo *Control Estadístico*.

Como se mencionó anteriormente este parámetro asume homocedasticidad, es decir, variabilidad constante en todo el rango de concentración. Al aplicar esta expresión, por ejemplo, a los datos de duplicados se obtiene el siguiente límite de repetibilidad:

$$\begin{aligned} r &= 2.8 \cdot s_r \\ r &= 2.8 \cdot 0.41 \\ r &= 1.15 \text{ mg/kg} \end{aligned}$$

Por lo tanto, para este ejemplo, la máxima diferencia tolerable en valor absoluto entre duplicados analizados en condiciones de repetibilidad es $r = 1.15$ mg/kg. Note que la expresión de cálculo de r es también aplicable a la desviación estándar de repetibilidad obtenida por el método de diseño experimental. El concepto es el mismo, sólo varía la metodología de estimación de s_r.

Este cálculo e interpretación son totalmente análogos si se quisiera obtener el límite de reproducibilidad:

$$r = 2.8 \cdot s_R$$

donde s_R es la desviación estándar de reproducibilidad.

Algunas observaciones de los estudios de precisión

29. Es un hecho conocido que en la mayoría de los métodos analíticos la precisión es función de la concentración (heterocedasticidad), por lo tanto, es muy importante evaluar esta condición dentro del rango de concentraciones que habitualmente tienen las muestras. En el ejemplo de análisis en duplicado propuesto en esta guía, advierta que el rango de concentración es estrecho [45 - 55] mg/kg, por lo tanto, no es sorpresivo que la variabilidad sea constante. Como se mencionó anteriormente, el método de Howarth- Thompson permite modelar s_r en función de la concentración, de esta manera el límite de repetibilidad también dependerá de la concentración, lo cual tiene sentido químico cuando el rango de concentraciones es amplio.

30. En la estimación de la precisión mediante un diseño estadístico, privilegie aumentar el número de niveles del factor que el número de replicados. Es decir, más días o más analistas con menores replicados permitirá obtener un estimador de precisión más consistente.

31. También es posible obtener la precisión de un método a partir de un estudio colaborativo entre varios laboratorios lo cual representa un diseño experimental más complejo y costoso, sin embargo, la información obtenida se acerca fielmente al verdadero desempeño del método cuando es aplicado en la rutina.

32. Los estimadores de precisión a bajas concentraciones, pero sobre el límite de cuantificación, podrían ser aparentemente altos en términos relativos (40 - 50 %). Sin embargo, esto es un hecho esperable y documentado en la literatura.

33. Si bien la guía ISO 5725 sugiere el límite de repetibilidad (reproducibilidad) como criterios de aceptación entre duplicados, la guía 657 de la Comunidad Europea propone un criterio basado en la ecuación de Horwitz. Este modelo, al igual que Howarth - Thompson, predice la desviación estándar de reproducibilidad en función de la concentración.

3. INCERTIDUMBRE DE MEDICIÓN

La Guía ISO 17025 exige a los laboratorios acreditados bajo esta norma, estimar la incertidumbre de sus mediciones, sin embargo, no describe ningún procedimiento específico para calcular este parámetro. Si bien existen varias formas de calcular la incertidumbre de medición que cumplen con el requerimiento de la ISO 17025 se recomienda, como una forma de armonizar el cálculo de incertidumbre entre los laboratorios autorizados por Sernapesca, evaluar este parámetro de acuerdo a la normativa internacionalmente aceptada, es decir, siguiendo las directrices de las Guías ISO/GUM y Eurachem.

Sin embargo, bajo argumentos técnicos, los cuales deberán ser entregados a Sernapesca para su evaluación, los laboratorios podrán estimar la incertidumbre de medición siguiendo las directrices de otras guías internacionalmente reconocidas, entre las cuales se pueden identificar:

34. ISO 21748:2017 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation.

35. ISO 11352:2012 Water quality -- Estimation of measurement uncertainty based on validation and quality control data

36. ISO/TS 21749:2005 Measurement uncertainty for metrological applications — Repeated measurements and nested experiments

37. ISO/TS 17503:2015 Statistical methods of uncertainty evaluation — Guidance on evaluation of uncertainty using two-factor crossed designs

Estas guías sugieren una aproximación del tipo *top-down* para estimar la incertidumbre, es decir, particionar la variabilidad observada (experimental) en fuentes individuales de incertidumbre. Para este fin, utiliza datos de la validación

(precisión), QAQC, ensayos de intercomparación, diseño de experimentos, etc. Esta aproximación es especialmente útil en metodologías del tipo *"caja negra"* o en métodos empíricos, en donde el mensurando está completamente definido por el método analítico (por ej: proteínas totales en harina de pescado mediante Kjeldhal).

La utilización de cualquiera de las aproximaciones está sujeta a criterios técnicos, estadísticos, económicos y la disponibilidad de datos históricos para el caso de la norma ISO 11352. En todos los casos, todas las aproximaciones entregan información válida y complementaria de la incertidumbre del método. Puede consultar en la bibliografía información adicional sobre la comparación entre los métodos de cálculo *bottom-up* (GUM) y *top-down*.

En la práctica, la incertidumbre puede ser el resultado del efecto de varias variables (controladas o no): instrumentos, pureza de los reactivos, condiciones ambientales, incertidumbre en el peso y volumen, diluciones, incertidumbre de los estándares de calibración, aproximaciones incorporadas al método de análisis, etc.

Esta guía comprende que evaluar el efecto de todos los posibles factores que podrían influenciar el resultado de una medición sería una tarea extensa y costosa. Por lo tanto, acogiendo el espíritu de la norma ISO 17025 el laboratorio deberá identificar, en base a su experiencia y/o datos de bibliografía, aquellos factores que tienen un efecto significativo.

Dentro del proceso de evaluación de la incertidumbre el laboratorio deberá evaluar las distintas contribuciones a la misma, teniendo en cuenta las posibles causas de incertidumbre agrupándolas bien por etapas (preparación de las muestras, extracción, preparación de las curvas de calibración, etc.) o por magnitudes implicadas (pesada, volumen, calidad de reactivos, sensibilidad de espectrofotómetros, etc.) Finalmente, se calculará la incertidumbre expandida mediante la combinación de todas las contribuciones consideradas.

Cada laboratorio deberá entregar a Sernapesca un informe que detalle cómo estima la incertidumbre de medición, el modelo de medición subyacente, los archivos de cálculo numérico-computacionales e indicar cuáles son los factores que más influyen en la incertidumbre final y las medidas tomadas para reducir su impacto. Aquellas contribuciones que no se usaron en los

cálculos finales porque se consideraron insignificantes deberán ser mencionadas.

Cálculo de incertidumbre basado en la guía GUM

La siguiente metodología está basada en la guía ISO/GUM, documento oficial e internacionalmente aceptado como guía básica para la estimación de la incertidumbre de medición. Además, se hace referencia a la guía Eurachem, la cual detalla y aplica los criterios de la guía ISO-GUM específicamente a Química Analítica. Ya que el cálculo de incertidumbre es un tema bastante extenso, sólo se hará una breve reseña de los detalles contenidos en estas guías y se mostrará un ejemplo de aplicación.

La estimación de la incertidumbre de medición involucra cinco etapas.

38. Especificación del modelo
39. Identificación de las fuentes de incertidumbre
40. Cuantificación de las incertidumbres estándar de las fuentes de error identificadas
41. Combinación de las incertidumbres estándar
42. Cálculo de la incertidumbre expandida

Especificación del Modelo

En esta etapa se debe establecer el mensurando y la relación funcional con las variables de entrada. Es decir, se debe declarar la ecuación de medición y todos los factores adicionales que estén involucrados en la obtención del valor del mensurando:

$$Y = f(X_1, X_2, \dots, X_n)$$

donde Y es el valor del mensurando, X_i son las variables de entrada y f es la relación funcional o ecuación de medición. Por ejemplo, en la determinación de mercurio (Hg) en choritos por ICP-MS (calibración estándar) una de las posibles relaciones funcionales entre concentración (Y) y las variables de entrada podría ser la siguiente:

$$C_{Hg} = \frac{C_{calib} \cdot V \cdot F_{dilucion}}{m}$$

en este caso el mensurando es $Y = C_{Hg}$, y las variables de entrada son: $X_1 = C_{calib}$ la concentración de la muestra interpolada en la curva de calibración, $X_2 = V$ es el volumen de aforo final, $X_3 = F_{dilucion}$ es un eventual factor de dilcuión y $X_4 = m$ m es la masa de la muestra.

Identificación de las fuentes de incertidumbre

Se debe listar todas las fuentes potenciales de incertidumbre que afecten el valor del mensurando. Se deben incluir aquellas que se deducen directamente de la ecuación de medición, así como las que tienen un efecto indirecto en el mensurando y que no están reflejadas en el modelo. En el ejemplo anterior, se pueden identificar las siguientes fuentes de incertidumbre:

- Incertidumbre de la balanza
- Incertidumbre de la calibración
- Incertidumbre del material de volumétrico (factor de dilución D)

Cuantificación de las incertidumbres estándar de las fuentes de error

identificadas

La guía ISO-GUM establece, en general, dos formas de estimar la incertidumbre de las fuentes de error: a través de métodos estadísticos (Tipo A) y mediante métodos no estadísticos (Tipo B). Se entregará una breve descripción de cada metodología, pues los detalles pueden ser consultados en la bibliografía adjunta:

43. Tipo A: Este método de evaluación de incertidumbre se basa en aplicar métodos estadísticos a una serie de mediciones repetidas. En general, si se dispone de una serie de n mediciones repetidas de un mensurando $\mathbf{X} = \{x_1, x_2, \dots x_n\}$, entonces la incertidumbre estándar de una *única* medición es equivalente a la desviación estándar de la serie de mediciones, es decir $u_x = s$. Si se utiliza el promedio de ls n mediones como valor del mensurando, entonces la incertidumbre estándar del promedio es $u_{\overline{X}} = s/\sqrt{n}$.

Ejemplo: para estimar la incertidumbre estándar de la lectura de absorbancia de un espectrofotómetro en la determinación de As u_{Abs}, se llevaron a cabo $n = 6$ lecturas independientes de una misma solución digerida y aforada de la muestra problema, los datos obtenidos son los siguientes:

Replicado	Absorbancia
1	0.405
2	0.415
3	0.400
4	0.412

5	0.406
6	0.410

El promedio de las lecturas es $\overline{Abs} = 0.408$ UA y la desviación estándar es $s = 0.005$ UA. Por lo tanto, la incertidumbre estándar de una única lectura de absorbancia es $u_{Abs} = s = 0.005$ UA. La incertidumbre estándar del promedio de las n lecturas es $u_{\overline{Abs}} = s/\sqrt{6} = 0.003$ UA.

3. Tipo B: Esta metodología utiliza la información provista por el fabricante, certificados de calibración, juicio experto, manuales de equipos, etc. Es necesario, sin embargo, convertir la información provista en incertidumbres estándar para poder combinarlas. En general, el fabricante u organismo certificador informará la incertidumbre expandida U, por lo tanto, la conversión a incertidumbre estándar se obtiene de $u = U/k$, donde k es establecido de acuerdo a la distribución de probabilidad elegida para evaluar la confiabilidad de la fuente de error. Las más utilizadas son la distribución Normal($k = 2$), Triangular($k = \sqrt{6}$) y Rectangular($k = \sqrt{3}$).

Ejemplo: El fabricante de un matraz de 100 mL clase A establece una tolerancia de $100 \pm 0{,}1$ mL. Por lo tanto, la incertiudmbre expandida corresponde a $U = 0{,}1$ mL. En general, para este tipo de fuentes de error se asume una ditribución triangular(o rectangular), por lo tanto, $k = \sqrt{6}$ y la incertidumbre estándar del volumen del matraz es $u = U/k = 0{,}1/\sqrt{6} = 0.04$ mL.

Combinación de las incertidumbres estándar

Para obtener, finalmente, la incertidumbre estándar del mensurando debemos combinar las incertidumbres de las fuentes de error previamente identificadas. El procedimiento de combinación sigue las directrices de la guía ISO-GUM la cual se basa en la Ley de Propagación de los Errores. Si Y es el mensurando el cual depende de las variables de entrada $X_1, X_2, \dots, X_n$ a través de una función f, entonces la incertidumbre estándar de Y u_Y es dada por:

$$u_Y = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial y}{X_i}\cdot u_{X_i}\right)^2 + \sum^{i\neq j}\frac{\partial y}{X_i}\cdot\frac{\partial y}{X_j}\cdot u_{X_i,X_j}}$$

donde $\frac{\partial y}{X_i}$ es la derivada parcial de Y con respecto a la variable X_i y u_{X_i,X_j} es la covarianza entre las variables X_i y X_j. Cuando las variables de entrada X_i son independientes, entonces las covarianzas $u_{X_i,X_j} = 0$ y la ecuación anterior se reduce a:

$$u_y = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial y}{X_i}\cdot u_{X_i}\right)^2}$$

Existen algunas reglas para operaciones aritméticas sencillas aplicando la ecuación anterior:

- Sumas y diferencias: Si el modelo involucra sólo sumas y/o diferencias $Y = X_1 \pm X_2 \pm \cdots \pm X_n$, entonces:

$$u_Y = \sqrt{u_{X_1}^2 + u_{X_2}^2 + \cdots + u_{X_n}^2}$$

Ejemplo: En la determinación de Hg en choritos por ICP-MS, la masa de la muestra problema puede (*y se recomienda*) ser obtenida por diferencia entre un vial vacío y el vial + muestra. En este caso, la ecuación de medición de la masa de la muestra es $m = m_{\text{vial + muestra}} - m_{\text{vial}}$, por lo tanto, aplicando la regla de la suma-diferencia la incertidumbre estándar de la masa de la muestra es u_m es:

$$u_m = \sqrt{u^2(m_{\text{vial + muestra}}) + u^2(m_{\text{vial}})}$$

- Productos y cocientes: Si el modelo involucra sólo productos y/o cocientes $Y = X_1 \times X_2 \times \ldots \div X_n$, entonces:

$$u_Y = Y \cdot \sqrt{\frac{u_{X_1}^2}{X_1^2} + \frac{u_{X_2}^2}{X_2^2} + \cdots + \frac{u_{X_n}^2}{X_n^2}}$$

Ejemplo: La incertidumbre estándar de un factor de dilución $F_{\text{dilucion}} = \text{V}_{\text{Final}} \div \text{V}_{\text{Inicial}}$ se obtiene a partir de:

$$u_{F_{\text{dilucion}}} = F_{\text{dilución}} \cdot \sqrt{\frac{u_{\text{V}_{\text{Final}}}^2}{\text{V}_{\text{Final}}^2} + \frac{u_{\text{V}_{\text{Inicial}}}^2}{\text{V}_{\text{Inicial}}^2}}$$

Cálculo de incertidumbre expandida

Una vez obtenida la incertidumbre estándar combinada u_Y es posible estimar la incertidumbre expandida U_Y a un nivel de confianza dado a través de la expresión:

$$U_Y = k \cdot u_Y$$

donde k es un factor de cobertura y depende del nivel de confianza deseado. En general, se asume una distribución Normal del mensurando por lo que $k = 2$. Con la incertidumbre expandida se puede establecer el intervalo de confianza para el valor del mensurando como $Y \pm U_Y$.

Los detalles sobre el cálculo de incertidumbre de los distintos métodos analíticos utilizados por los laboratorios de la Red están fuera del alcance de

esta guía, sin embargo, pueden consultarse en la bibliografía adjunta. Se desarrollará un ejemplo práctico con el fin de estandarizar la estimación de este parámetro.

Ejemplo de Cálculo de Incertidumbre Método ISO-GUM. Método de determinación de Hg en choritos por ICP-MS.

44. Especificación del modelo

El mensurando es la concentración de mercurio total en choritos C_{Hg} μg/kg. La ecuación de medición es la siguiente:

$$C_{Hg} = \frac{C_{calib} \cdot V \cdot F_{dilucion}}{m}$$

donde $y = C_{Hg}$, y las variables de entrada son: $X_1 = C_{calib}$ la concentración de la muestra interpolada en la curva de calibración, $X_2 = V$ es el volumen de aforo final, $X_3 = F_{dilucion}$ es un eventual factor de dilcuión y $X_4 = m$ m es la masa de la muestra.

Para desarrollar este ejemplo se asumirá la que la masa de de la muestra es de $m = 0.5000$ g y el volumen de aforo es 50 mL. Los calibrantes preparados volumétricamente en matriz ácida corresponden a 0 - 1 - 2 - 5 - 10 ng/mL (ppb).

4. Identificación de las fuentes de incertidumbre

De acuerdo a la ecuación de medición, se identifican las siguientes fuentes de incertidumbre:

- $u_{C_{calib}}$ Incertidumbre de la curva de calibración
- u_V Incertidumbre del material volumétrico (matraz 50 mL)
- u_m Incertidumbre de la balanza al medir la masa de la muestra

Por supuesto que existen otras fuentes de incertidumbre, sin embargo, el método GUM no las incorpora explícitamente. Por ejemplo, en este modelo de medición no está explícita la incertidumbre de los calibrantes, tal como veremos A continuación,.

Por simplicidad asumiremos que no existe dilución de la muestra antes de la lectura instrumental, por lo tanto, $F_{dilucion} = 1$ y $u_{F_{dilucion}} = 0$.

6. Cuantificación de la las incertidumbres estándar de las fuentes de error identificadas

7. $u_{C_{calib}}$ Incertidumbre de la curva de calibración

Para estimar la incertidumbre de la concentración interpolada de la solución digestada de la muestra en la curva de calibración, utilizaremos la expresión indicada en la guía Eurachem. Esta metodología es aplicable a cualquier método analítico que determina la concentración de la muestra problema a través de la interpolación de la señal instrumental de la muestra en una curva de calibración lineal de la forma:

$$y = \beta_0 + \beta_1 \cdot x + \epsilon$$

donde y es la señal instrumental (ej: absorbancia, áreas, volts, etc.), x es la concentración, β_0 es un intercepto, β_1 es la pendiente de la curva de calibración y ϵ es un error aleatorio que se asume tiene una distribución Normal con media 0 y varianza σ^2. Este modelo de calibración asume como válidas tres suposiciones:

- La única fuente de error proviene de la señal instrumental y, es decir, los calibrantes x no tienen incertidumbre de preparación. Desde el punto de vista metrológico esta suposición es falsa, ya que los calibrantes si poseen incertidumbre de preparación debido a fuentes de error provenientes, por ejemplo, de la balanza, del material volumétrico utilizado o del certificado de pureza del material de partida. Desde el punto de vista estadístico existen métodos de regresión que consideran *errores en ambos ejes* lo cual permite incorporar explícitamente la incertidumbre de los estándares, sin embargo, estos métodos se encuentran fuera del alcance de esta guía y pueden consultarse en la bibliografía.

- La varianza del error aleatorio ϵ es la misma en todo el rango lineal (propiedad de *homocedasticidad*). Es decir, el error de la señal instrumental es el mismo en todo el rango lineal de concentraciones. Si bien en rangos lineales de 2 órdenes de magnitud esta suposición es plausible, es necesario demostrar la homocedasticidad a través de un test estadístico formal (Test de Durbin-Watson) o mediante pruebas gráficas de análisis de residuos, técnicas que están fuera del alcance de esta guía.

- Los errores aleatorios ϵ son independientes, es decir, la medición de una muestra no influye en la medición de las muestras sub-siguientes. Aunque una adecuada aleatorización de las mediciones analíticas disminuiría algún tipo de sesgo no químico, se reconoce que algunas técnicas analíticas tienen un marcado *efecto memoria* lo cual implica que las mediciones tienen algún grado de correlación, especialmente aquellas cuyos instrumentos poseen *autosampler*. Se sugiere una adecuada limpieza del sistema entre inyecciones.

Por lo tanto, bajo este modelo la incertidumbre de calibración representa *solamente* fuentes de errores instrumentales y no incorpora la incertidumbre de los calibrantes ni la del material de referencia utilizado en la preparación de éstos.

Se asume que el modelo lineal de calibración ha sido evaluado según las directrices de indicadas en la sección *Linealidad* y se ha establecido el rango lineal de calibración para el método analítico. De acuerdo a la ecuación de calibración, si y_0 es la señal instrumental de la muestra problema, entonces la concentración interpolada de la muestra C_{calib} se obtiene según:

$$C_{calib} = \frac{y_0 - \beta_0}{\beta_1}$$

Lo que corresponde, entonces, es aplicar las directrices de ISO/GUM a C_{calib} para obtener su incertidumbre estándar, aplicando primero en el numerador la regla de las diferencias y luego la regla de un cociente. Sin embargo, se debe tener en consideración algunos conceptos estadísticos adicionales:

- Como se describió anteriormente, la única fuente de incertidumbre del modelo de calibración lineal es la señal instrumental y. Este hecho tiene como consecuencia que al estimar la pendiente y el intercepto, los estimadores de estos parámetros tengan una incertidumbre asociada.

- Además, cuando se aplica el método de los mínimos cuadrados, la covarianza entre pendiente e intercepto no es nula, por lo tanto, debe ser considerada en el cálculo.

En resumen, al calcular la concentración de la muestra problema de acuerdo a la ecuación de calibración existen cuatro fuentes de incertidumbre:

- Pendiente $u_{\hat{\beta}_1}$
- Intercepto $u_{\hat{\beta}_0}$
- Covarianza entre pendiente e intercepto $u_{\hat{\beta}_0,\hat{\beta}_1}$
- Señal instrumental de la muestra problema u_{y_0}

Afortunadamente, existe una aproximación muy buena para estimar $u_{C_{calib}}$, la cual se cumple en todas las calibraciones en química analítica:

$$u_{C_{calib}} = \frac{\hat{\sigma}_{y/x}}{\hat{\beta}_1}\sqrt{\frac{1}{n} + \frac{1}{m_0} + \frac{(C_{calib} - \overline{C})^2}{\sum(C_i - \overline{C})^2}}$$

donde $\hat{\sigma}_{y/x}$ es el error estándar de la regresión, β_1 es la pendiente, n es el número de calibrantes, m_0 es el número de replicados independientes de la muestra problema, C_{calib} es la concentración de la muestra interpolada en la curva de

calibración, $\overline{C}$ es el promedio de las concentraciones de los calibrantes y C_i representa la concentración del calibrante i.

Diseño experimental de la curva para minimizar la incertidumbre de calibración

Al observar la ecuación de incertidumbre de calibración se puede deducir aquellos factores que el laboratorio puede modificar y/o establecer para minimizar la incertidumbre de calibración:

45. Aumentar el número de calibrantes n, disminuyendo así los factores $1/n$ y $\frac{(C_{calib}-\overline{C})^2}{\sum(C_i-\overline{C})^2}$.

46. Aumentar el número de replicados independientes de la muestra problema m_0, disminuyendo así el término $1/m_0$.

47. Maximizar el uso del rango lineal, esto aumentará el término $\sum(C_i-\overline{C})^2$.

48. Diseñando la curva de calibración de tal manera que la concentración de la muestra problema C_{calib} se encuentre en la zona media de la curva de calibración $\overline{C}$. Así, cuando $C_{calib}=\overline{C}$ entonces $\frac{(C_{calib}-\overline{C})^2}{\sum(C_i-\overline{C})^2}=0$. Esto queda en evidencia en la siguiente figura, la cual muestra una curva de calibración y las bandas de incertidumbre asociadas, la cual alcanza un mínimo en el centro de a curva.

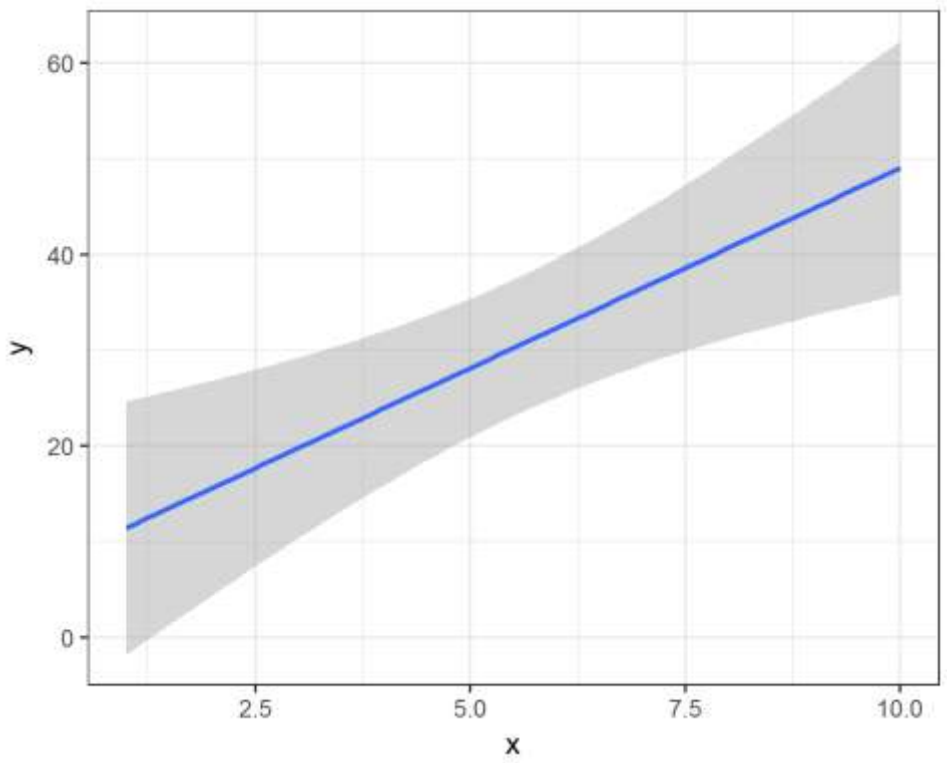

Figura 9. Incertidumbre de calibración

Adicionalmente, un buen diseño de curva de calibración debe ser equiespaciado, es decir, por ejemplo: 10-20-30-40-50 ppm, considerando al menos 5 niveles de concentración $n = 5$. Notar que un aumento de n y m_0 no necesariamente conlleva a una disminución significativa de la incertidumbre de calibración desde el punto de vista práctico, lo cual podría significar costos innecesarios en tiempo y recursos.

Si el sistema de medición se encuentra en bajo Control Estadístico, específicamente los parámetros de pendiente β_1 y desviación estándar de la regresión $\sigma_{y/x}$, entonces es posible calcular aproximadamente el número de replicados independientes de la muestra problema necesarios para alcanzar un cierto nivel de incertidumbre. Bajo el supuesto que todos los demás términos de la ecuación de incertdiumbre de calibración $u_{C_{calib}}$ permanezcan constantes. Por ejemplo, si se desea obtener una incertidumbre de calibración del 5 % ($u_{C_{calib}} = 0.05 \cdot C_{calib}$), entonces es factible despejar m_0 de la ecuación $u_{C_{calib}}$:

$$\frac{1}{m_0} = \left(\frac{\beta \cdot u_{C_{calib}}}{\sigma_{y/x}}\right)^{-1} - \frac{1}{n} - \frac{(C_{calib} - \overline{C})^2}{\sum(C_i - \overline{C})^2}$$

Esta expresión es bastante útil considerando que para muchos métodos analíticos, la curva de calibración es un de los factores con mayor aporte a la incertidumbre final.

Continuando con el ejemplo de la determinación de Hg en choritos por ICP-MS, se muestran A continuación, los datos, análisis estadístico y gráficos de la calibración:

Tabla 7. Datos de calibración Hg ng/mL

Concentración ng/mL	Cuentas
0	33
1	509
2	1041
5	2590
10	5172

Tabla 8. Análisis estadístico de la regresión

Término	Coeficiente	Error std	t-Student	p-value
Intercepto	13	10.5	1.2	0.32
Pendiente	516	2.1	249.7	0.00

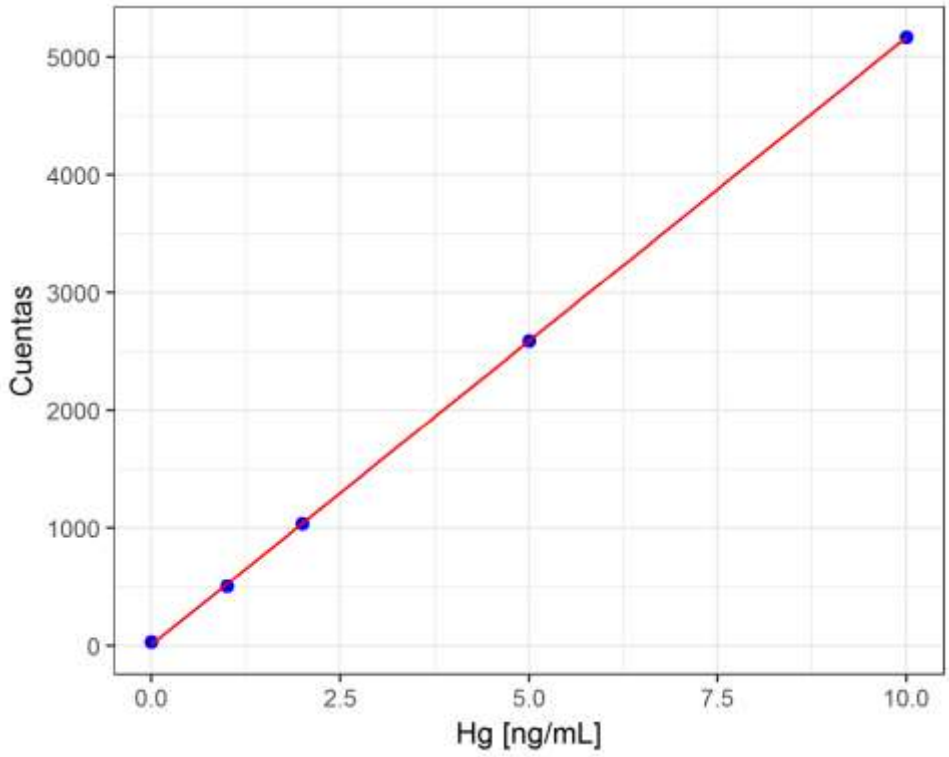

Figura 10. Curva de calibración Hg por ICP-MS

Si la muestra problema tiene una señal instrumental de $Y_0 = 850$ cuentas, entonces, la concentración interpolada en la curva de calibración es $C_{calib} = 1.62$ ng/mL y su incertidumbre estándar es $u_{C_{calib}} = 0.04$ ng/mL.

5. u_V Incertidumbre del volumen de aforo

La incertidumbre estándar del volumen del aforo $V = 50$ mL es obtenida a partir de los datos de incertidumbre del material volumétrico utilizado, en este caso, del matraz. La maginitud de este componente se calcula a partir de la combinación de tres fuentes:

- u_{tol} La incertidumbre informada por el fabricante en el certificado original del lote de producción o aquella impresa en el material de vidrio. Esta incertidumbre es de Tipo B.

- u_{rep} Incertidumbre debido a la repetibilidad del analista en el llenado de un mismo volumen. Esta incertidumbre (Tipo A) es evaluada mediante métodos estadísticos a través de la replicación (n) del vertido de un volumen fijo de agua desionizada por un único analista, y el registro gravimétrico de la cantidad de agua. La incertidumbre estándar se expresa como la desviación estándar de los n replicados.

- u_{temp} La incertidumbre debido a la diferencia de temperatura registrada durante la calibración del material de vidrio y aquella que se registra durante los análisis de rutina. Esta incertidumbre expandida es expresada por la siguiente ecuación:

$$U_{\text{temp}} = V \cdot \gamma_{\text{H}_2\text{O}} \cdot \Delta T$$

donde V es el volumen certificado del material de vidrio, $\gamma_{\text{H}_2\text{O}} = 2.1 \cdot 10^{-4}°\text{C}^{-1}$ es el coeficiente de expansión del agua y ΔT es la variabilidad de la temperatura de trabajo del laboratorio en relación a la temperatura de calibración del material. Finalmente, se obtiene asume una distribución rectangular de la variable *Temperatura* obteniéndose una incertidumbre estándar $u_{\text{temp}} = U_{\text{temp}}/\sqrt{3}$ mL.

Esta expresión debe utilizarse cuando exista un certificado de calibración que informe la temperatura de referencia registrada durante el proceso de certificación el cual puede provenir del mismo fabricante o de alguna institución metrológica nacional que certifique este tipo material. Generalmente la temperatura de calibración se especifica a 20°C.
La incertidumbre combinada estándar del volumen del aforo se obtiene a través de la siguiente expresión:

$$u_{\text{V}} = \sqrt{u_{\text{tol}}^2 + u_{\text{rep}}^2 + u_{\text{temp}}^2}$$

Continuando con el ejemplo, para estimar esta incertidumbre se calculó la incertidumbre de los tres componentes que contribuyen al error del volumen del aforo:

- u_{tol}: El fabricante informa una incertidumbre del matraz de ± 0.05 mL, por lo tanto, al asumir una distribución triangular se obtiene una incertidumbre estándar de $u_{\text{tol}} = 0.05/\sqrt{6} = 0.02$ mL.

- u_{rep}: Se llevaron a cabo $n = 6$ replicados de aforo de 50 mL con agua desioniada en un matraz utilizado en el trabajo de rutina del laboratorio, registrando la masa de agua en condiciones de repetibilidad (un único analista). La siguiente tabla muestra los valores obtenidos en mL, es decir, transformados desde unidades de masa [g] utilizando una tabla de densidades del agua y ajustando por la temperatura (se incluyen 4 decimales). La incertidumbre de repetibilidad de llenado es $u_{\text{rep}} = 0.01$ mL.

Tabla 9. Repetibilidad llenado matraz 50 ± 0,05 mL

Replicado	Volumen [mL]
1	50.4778
2	50.4515
3	50.4617
4	50.4732
5	50.4475
6	50.4583

- u_{temp}: Para calcular la incertidumbre de este componente se estimó que la máxima diferencia entre la temperatura de calibración indicada por el fabricante (20°C) y la temperatura ambiental del laboratorio en condiciones de rutina es $\Delta T = \pm 4°\text{C}$. Con esta información la incertidumbre expandida del efecto de la temperatura es dada por:

$$\begin{aligned} U_{\text{temp}} &= V \cdot \gamma_{\text{H}_2\text{O}} \cdot \Delta T \\ &= 50 \cdot 2.1 \cdot 10^{-4} \cdot 4 \\ &= 0.04 \text{ mL} \end{aligned}$$

Finalmente, se asume una distribución rectangular de la variable *Temperatura* obteniéndose una incertidumbre estándar $u_{\text{temp}} = 0.02$ mL.

Al combinar las tres fuentes de incertidumbre se obtiene una incertidumbre estándar del volumen del matraz de:

$$\begin{aligned} u_V &= \sqrt{u_{\text{tol}}^2 + u_{\text{rep}}^2 + u_{\text{temp}}^2} \\ &= \sqrt{4.17 \times 10^{-4} + 1.42 \times 10^{-4} + 5.88 \times 10^{-4}} \\ &= 0.03 \text{ mL} \end{aligned}$$

- u_{m}: La masa de la muestra problema es 0,5000 g. La incertidumbre de la masa de la muestra se obtiene directamente del certificado de calibración de la balanza. Para este ejemplo, el certificado informa una incertidumbre expandida para un nivel nominal de 0,5000 g de $U_{\text{m}} = 0.0001$ g con $k = 2$, por lo tanto, la incertidumbre estándar se obtiene a partir de $u_{\text{m}} = U_{\text{m}}/2 = 0{,}00005$ g.

6. Cálculo de la incertidumbre combinada:

La siguiente tabla muestra el *budget* total de incertidumbre para el modelo de medición:

Tabla 10. Componentes de incertidumbre

Variable	Valor	u.std
Curva calibración	1.6	0.04
Volumen aforo	50.0	0.03
masa muestra	0.5	0.00
C_{Hg}	162.4	3.63

Con esta información consolidada, ahora es posible obtener la incertidumbre combinada en base a la propagación de errores de la ecuación de medición:

$$C_{Hg} = \frac{C_{calib} \cdot V}{m}$$

$$u_{C_{Hg}} = C_{Hg} \sqrt{\left(\frac{u_{C_{calib}}}{C_{calib}}\right)^2 + \left(\frac{u_V}{V}\right)^2 + \left(\frac{u_m}{m}\right)^2}$$

$$u_{C_{Hg}} = 3.6$$

Finalmente, la incertidumbre expandida $U_{C_{Hg}}$ es obtenida a través de de la siguiente expresión:

$$\begin{aligned} U_{C_{Hg}} &= k \cdot u_{C_{Hg}} \\ &= 2 \cdot u_{C_{Hg}} \\ &= 2 \cdot 3.63 \\ &= 7.26 \text{ ng/g} \end{aligned}$$

donde $k = 2$ es el factor de cobertura para un 95% de confianza asumiendo que la concentración de Hg obtenida por la ecuación de medición puede ser modelada razonablemente mediante una distribución Normal. Por lo tanto, considerando las cifras significativas, el resultado de la concentración de Hg en esta muestra de choritos se debiera informar como 162 ± 7 ng/g, con un 95% de confianza ($k = 2$).

A continuación, la figura muestra la contribución de cada uno de los factores que afectan la incertidumbre de la concentración final de Hg, en ella se puede apreciar que el factor que más aporta a la incertidumbre es la curva de calibración. La balanza y el material volumétrico, como es sabido, no influye significativamente en el *budget* de incertidumbre de los métodos instrumentales. Por lo tanto, esta información es valiosa pues presenta un indicio hacia dónde deben dirigirse los esfuerzos si se desea disminuir la incertidumbre de medición. Esta es la gran ventaja del método GUM frente a los métodos *top-down* basados en datos de validación o de QAQC.

Note que para describir la contribución de cada factor, se debe graficar el producto entre el coeficiente de sensibilidad por la incertidumbre estándar correspondiente, es decir, $\partial(y)/\partial(x) \cdot u_x$.

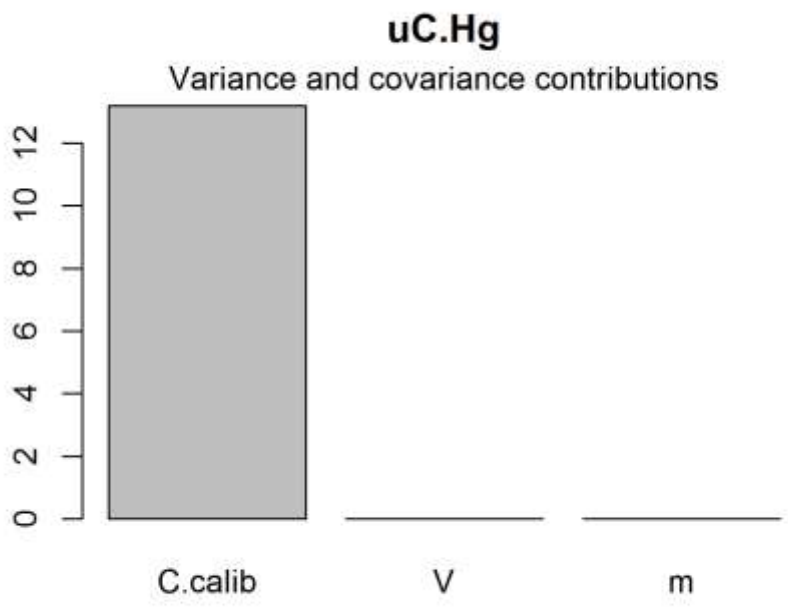

Figura 11. Contribución de cada factor a la incertidumbre total

4. LÍMITES DE DETECCIÓN Y CUANTIFICACIÓN

Esta guía describe protocolos específicos para la estimación del límite de detección (LoD) y Cuantificación (LoQ) de acuerdo a la técnica analítica utilizada. La elección de la metodología estadística dependerá entre otros factores de las características del método analítico y de la disponibilidad de muestras blanco. Esta guía sólo aborda la estimación de LoD y LoQ de métodos instrumentales.

Idealmente, estos parámetros debieran obtenerse a partir del análisis de blancos analíticos, es decir, toda la matriz excepto el analito de interés. En el caso del análisis de metales, este tipo de muestra es prácticamente imposible de obtener (al menos para aquellos elementos que están normados a nivel internacional). En el caso de mensurandos orgánicos (pesticidas, antibióticos, PCBs, etc.) si bien es posible conseguir muestras de matriz con ausencia de estos compuestos (o en cantidades no detectables), en la práctica la señal transiente de la cromatografía (en función del tiempo) no permite obtener una señal para una muestra blanco en términos de área.

No tiene sentido químico integrar ruido instrumental o línea base en el tiempo de retención del analito en una muestra blanco.

Utilizar blancos reactivos sólo permite la estimación de un límite de detección instrumental. Para métodos espectrofotométricos o ICP-MS este límite instrumental puede estimarse mediante blancos reactivos, pues éstos poseen una señal instrumental distinta de cero. Sin embargo, en métodos cromatográficos esto no es factible, por lo que se debe recurrir a otros métodos como por ejemplo la relación señal ruido S/N que se muestra en la siguiente figura. Este método está basado en el procesamiento computacional de las señales registradas en un cromatograma de tal forma de cuantificar la razón señal-ruido de las alturas de peak, es decir, basado en la señal electrónica que registra el equipo y no en las determinaciones de áreas.

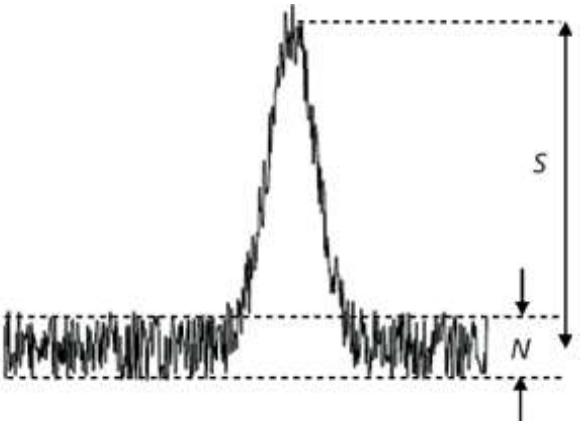

Figura 12. Razón señal/ruido cromatográfica

En general, para la determinación de Lod y LoQ es necesario disponer de i) muestras blanco, es decir, aquellas que corresponden a la matriz a analizar excepto el mensurando (o en cantidades no detectables), o bien, ii) muestras que tienen una muy baja concentración, cercanas o debajo al límite de cuantificación esperable, el cual puede ser obtenido de revisión bibliográfica.

Para estimar ambos parámetros se debe preparar $n = 10$ muestras blanco, o muestras con bajos niveles de concentración (cromatografía), y someterlas a todo el proceso analítico.

El límite de detección (LoD) es obtenido de la siguiente expresión:

$$\mathrm{LoD} = 3 \cdot s_{\mathrm{blanco}}$$

Donde s_{blancos} es la desviación estándar de las concentraciones de los $n = 10$ blancos analizados, expresadas en las unidades de concentración de la muestra problema (por ejemplo: mg/kg).

El límite de cuantificación se estima a partir de la siguiente expresión:

$$\text{LoQ} = 10 \cdot s_{\text{blancos}}$$

5. SESGO Y RECUPERACIÓN

El Vocabulario Internacional de Metrología (ISO/VIM) define sesgo como:

2.18 sesgo de medida, m sesgo, m *valor estimado de un error sistemático*
2.17 (3.14) error sistemático de medida, m error sistemático, m *componente del error de medida que, en mediciones repetidas, permanece constante o varía de manera predecible*
NOTA 1 El valor de referencia para un error sistemático es un valor verdadero, un valor medido de un patrón cuya incertidumbre de medida es despreciable, o un valor convencional de una magnitud.
Por lo tanto, para estimar el sesgo de un método analítico es necesario disponer de un *valor verdadero* de concentración de una muestra. Desde el punto de vista metrológico esto es imposible de obtener, por lo que en la práctica se utiliza un material de referencia certificado cuyo valor se asume como *verdadero*. Por lo tanto, hay que tener en consideración el siguiente hecho:
Todos los métodos analíticos tienen sesgo, el problema se reduce a evaluar si ese sesgo es significativo para los propósitos que se utilizarán los datos analíticos generados en el laboratorio.
Actualmente no existen materiales de referencia certificados para todas las matrices y analitos de interés hidrobiológico. por lo tanto, para estimar la veracidad de un método analítico se distinguen dos situaciones:
Existencia de un Material de Referencia Certificado (MRC)

Cuando el laboratorio utiliza un MRC para evaluar la veracidad del método analítico el procedimiento sugerido es analizar al menos $n = 6$ replicados independientes del material en condiciones de repetibilidad y comparar el

resultado obtenido con los datos provenientes del certificado. En este caso, al utilizar un MRC conceptualmente se estima el *sesgo* del método. El siguiente protocolo permite evaluar si existe un sesgo metrológicamente significativo y complementar esta información con los requerimientos del cliente y/o del sistema de QA/QC:

49. Se debe estimar el sesgo b y el estadístico Δ_c a partir de las siguientes expresiones:

$$b = \overline{x} - x_0$$

donde $\overline{x}$ y x_0 corresponden al promedio de los $n = 6$ replicados del MRC y a la concentración certificada del material, respectivamente.

$$\Delta_c = t_{n-1,1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} + U_{\mathrm{MRC}}$$

donde σ es la desviación estándar de los $n = 6$ replicados, U_{MRC} corresponde a la incertidumbre expandida del material de referencia certificado, la cual puede ser obtenida desde el certificado del material y $t_{n-1,1-\alpha/2}$ es la variable T Student con $n - 1$ grados de libertad y un $palpha = 0.05$.

6. Una vez calculado el sesgo y su incertidumbre, se lleva a cabo el siguiente test estadístico:

$$\text{Evaluación de sesgo} = \begin{cases} |b| \leq \Delta_c & \text{No hay evidencias de la presencia de sesgo} \\ |b| > \Delta_c & \text{Existe un sesgo estadísticamente significativo} \end{cases}$$

Observaciones e interpretación del test estadístico

- Resultado del Test: No hay evidencias de la presencia de sesgo El test propuesto considera ambas fuentes de incertidumbre: la precisión del laboratorio σ y la incertidumbre del material de referencia U_{MRC}. Si el resultado del test indicara que no hay evidencia de sesgo en el método analítico, esto NO significa ausencia sesgo, sino que debe interpretarse que, con los datos disponibles, no ha sido posible encontrar diferencias significativas con el valor de referencia. (*Ausencia de evidencia, no es evidencia de ausencia*)

- Resultado del Test: Existe un sesgo estadísticamente significativo

Si el resultado del test arrojara la conclusión que existe un sesgo estadísticamente significativo, esta información se debe analizar e interpretar en conjunto con los requerimientos del sistema de QA/QC y del cliente. Es decir, no es un dato aislado. Lo que se descubrió es que el laboratorio tiene una diferencia metrológica con el valor de certificado del material. ¿Es realmente importante esa diferencia?¿Tiene

relevancia práctica? Desde el punto de vista de la medición, todos los métodos analíticos son sesgados, por lo tanto, lo que el laboratorio debe evaluar es si el sesgo obtenido cumple con el propósito del método y requerimientos del cliente. Para lo cual el laboratorio debe establecer previamente cuál será el intervalo de tolerancia aceptable (Δ) para el sesgo del método, de tal forma de comparar el sesgo experimentalmente obtenido con la tolerancia preestablecida. La regla de decisión general es la siguiente:

- si el valor promedio del sesgo está dentro de los límites de tolerancia, implica que el método satisface los requerimientos de QA/QC.
- si el valor promedio del sesgo está fuera de los límites de tolerancia, implica que el método no satisface los requerimientos. Es decir, se ha detectado un sesgo que, además, no cumple con las tolerancias preestablecidas.

Si no existieran las tolerancias preestablecidas puede considerar, al menos inicialmente, el Δ_c como una primera aproximación y construir cartas control de sesgo utilizando el Δ_c como límite y cero como valor objetivo. Esto es especialmente cierto cuando el material de referencia ha sido certificado por un Instituto Nacional de Metrología (NIST, LGC, KRISS, etc.) con un método de la más alta jerarquía metrológica, pues el valor certificado tendrá una incertidumbre muy baja, imposible de lograr por un laboratorio de ensayo.

Por otro lado, actualmente existen en el mercado materiales de referencia certificados cuya incertidumbre es bastante grande para algunos mensurandos, lo que genera una suerte de "permisividad" en una prueba de sesgo. Por ejemplo, el material DOLT-5 (metales en *Dog fish*) tiene un valor certificado de Hg de 0.44 ± 0.18 mg/kg, es decir, un 40% de incertidumbre. En estos casos se sugiere ue el test de sesgo sólo considere la incertidumbre del material, es decir, no incluya la precisión del laboratorio. Por lo tanto, el sesgo b debe compararse directamente con la incertidumbre expandida del material U.

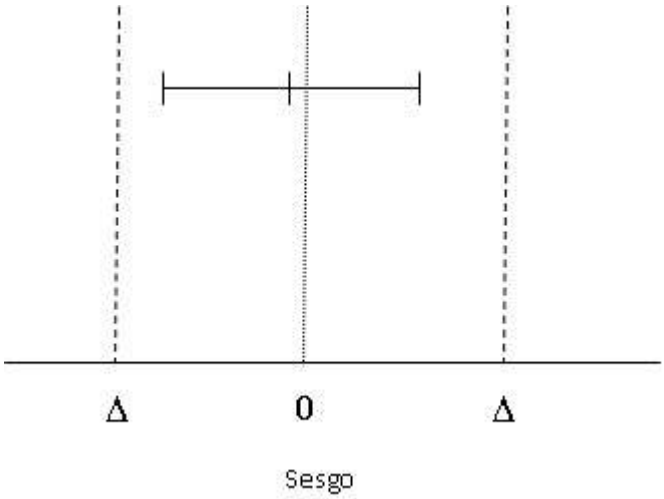

Figura 13. El sesgo está dentro de los límites de tolerancia permitidos

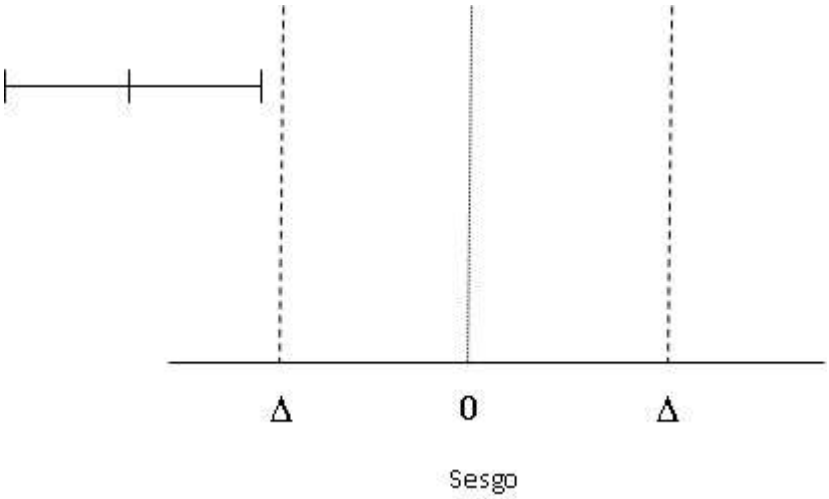

Figura 14. El sesgo está fuera de los límites de tolerancia permitidos

Recuperación: Uso de muestra fortificada

En esta situación no se aplica, rigurosamente, el concepto de sesgo, sino más bien se estima la recuperación del método, para lo cual existen varias aproximaciones. Se sugiere, como procedimiento general, estimar la recuperación a partir de $n = 6$ blancos de matriz fortificados, los cuales son sometidos a todo el proceso analítico. La recuperación es calculada a partir de la expresión $R = \mathrm{C}_{\mathrm{experimental}}/\mathrm{C}_{\mathrm{teorica}} \cdot 100$ y su incertidumbre puede estimarse a partir de la repetibilidad, es decir,

$u(R) = \sigma_R/\sqrt{n}$. Al igual que en el caso del sesgo, el laboratorio puede evaluar si la recuperación obtenida cumple con los requerimientos internos de QA/QC o aquellos requeridos por el cliente.

Esta guía está alineada con las directrices de la guía EC/657/2002 en relación a los niveles de concentración en que debe evaluarse la recuperación de un método analítico:

50. Cuando existe un Límite Máximo de Residuo (*MRL*) la evaluación de la recuperación se llevará a cabo en 0.5, 1 y 1.5 veces el *MRL*. El laboratorio deberá tener en cuenta los múltiples MRL definidos por cada uno de los mercados de destino para aplicar esta directriz.
51. Cuando se ha establecido un Límite de Funcionamiento Mínimo Exigido (*MRPL*) se evaluará la recuperación en 1, 1,5 y 2 veces el *MRPL*.

Para evaluar la recuperación del método se calcula el siguiente test estadístico:

$$t_{\mathrm{R}} = \frac{\overline{\mathrm{R}} - 100}{\sigma_{\mathrm{R}}/\sqrt{n}}$$

donde $\overline{\mathrm{R}}$ y σ_{R} corresponden la recuperación promedio y desviación estándar de los $n = 6$ blancos fortificados, respectivamente. Este estadístico calculado es comparado con un T de Student con $n - 1$ grados de libertad y $\alpha = 0.05$. La regla de decisión es la siguiente:

Evaluación de recuperación

$$= \begin{cases} |\mathrm{R}| \leq {}_{-1,=0.05} & \text{No hay diferencias significativas con 100\% Recuperación} \\ |\mathrm{R}| > {}_{-1,=0.05} & \text{Existen diferencias significativas con 100\% Recuperación} \end{cases}$$

Análogamente a la regla de decisión para el sesgo, los resultados del test de recuperación deben evaluarse en conjunto con los requerimientos de QA/QC y del cliente. Evidencia de diferencias significativas con el 100% de recuperación no implican necesariamente un trabajo analítico no conforme.

Observaciones a los estudios de recuperación:

52. Los estudios de recuperación con muestras fortificadas tienen la desventaja que, a menos que el analito se encuentre en forma libre en la matriz, no simula completamente las interacciones de éste con los otros constituyentes de la muestra, por ejemplo, unión a proteínas o reacciones de metabolización.

53. En los casos en que no se disponga de un blanco analítico en matriz, se puede utilizar una muestra incurrida fortificada, pero considerando en los cálculos de recuperación la concentración original de la muestra.

54. Se recomienda utilizar un estándar de fortificación de distinto origen (marca, lote, etc.) que el utilizado para preparar la curva de calibración. El argumento metrológico que sustenta esta directriz se basa en el concepto de independencia entre la calibración y la muestra control (fortificación), entendiendo como *independencia* aquellas actividades o diseños experimentales que minimicen las fuentes comunes de sesgos entre un calibrante y un control. Por ejemplo, no es correcto verificar la preparación de una curva de calibración prediciendo la concentración de uno de los mismos calibrates pues siempre será una prueba conforme. El mismo concepto metrológico aplica a una prueba de recuperación.

6. ROBUSTEZ

Los estudios de robustez se proponen evaluar los efectos sobre el desempeño del método de variaciones de uno o varios factores, cuanti y/o cualitativos, propios de una metodología analítica específica.

Estudiar todos los posibles factores que podrían tener un efecto significativo en el desempeño de la metodología es un desarrollo extenso en el tiempo y que consume muchos recursos, tanto monetarios como humanos. Por esta razón, el principal objetivo de un estudio de robustez es la eficiencia: obtener la mayor cantidad de información posible, utilizando el mínimo de recursos. Para cumplir este objetivo, es necesario aplicar los conceptos de Diseño Estadístico de Experimentos. Este tema es muy amplio para ser descrito en forma detallada en esta guía de validación, por lo tanto, se presentará el esquema general de un diseño experimental y se ejemplificará con algunos casos. Se recomienda la lectura de las guías de Validación de la Comunidad Europea EC/657/2002 y Guía Eurachem para ejemplos adicionales.

Esta guía se focalizará en los métodos propuestos por la guía 657 denominado Diseños Plackett-Burman.

Los pasos a seguir en un estudio de robustez son los siguientes:

55. Identificación de los factores que serán estudiados
56. Definición de los distintos niveles para los factores
57. Selección del diseño experimental
58. Definición del protocolo experimental
59. Definición de la respuesta que será utilizada como variable dependiente
60. Ejecución de los experimentos

61. Cálculo de los efectos
62. Análisis estadístico de los efectos
63. Obtención de conclusiones químicamente relevantes para mantener el método analítico bajo control estadístico.

Esquema general de un Diseño Experimental de Robustez Plackett-Burman

La herramienta estadística básica para evaluar la robustez de un método analítico son los diseños experimentales factoriales. Estos diseños permiten no tan solo evaluar el efecto de un factor (*efecto principal*) sino que también entrega valiosa información acerca de la interacciones entre los factores: ¿Existe alguna interacción entre pH y Temperatura en una reacción de derivatización post-columna, para un método particular de HPLC-Fluorescencia?

El problema de los diseños factoriales es que el número de experimentos crece rápidamente cuando aumenta el número de factores a estudiar. Es por ello que la mayoría de las guías de validación proponen un diseño experimental de *screening* con el fin de investigar posteriormente, sólo aquellos factores que tienen un efecto significativo en el desempeño del método.

Los diseños Plackett-Burman permiten estudiar factores con $+1$ experimentos y son ideales para evaluar muchos factores con pocos experimentos, donde es un múltiplo de 4. Sin embargo, esta simplicidad tiene un costo, pues no es posible detectar interacciones entre los factores experimentales, por lo tanto, sólo es posible obtener los *efectos principales*. Desde el punto de vista del estudio de robustez se ajustan muy bien al objetivo de eficiencia.

Esta guía no pretende entregar un conocimiento detallado de todos los diseños experimentales PB que podrían ser aplicables al estudio de robustez, sino que se propone ejemplificar su aplicación y detallar los cálculos necesarios. Se puede consultar la bibliografía adjunta para encontrar ejemplos desarrollados en química analítica.

Ejemplo de estudio de robustez. Determinación de cloranfenicol en salmón

por *Negative Chemical Ioniation GC-MS*

En primer lugar hay que establecer cuáles serán los factores a estudiar y los niveles o rangos que modificaremos del método original. En este ejemplo, el

método indica las siguiente condiciones para los siguientes factores experimentales:

64. Reactivo Derivatizante: BSFTA
65. Tiempo de reacción de derivatización: 40 min
66. Cantidad de reactivo: *100*
67. Temperatura de derivatización: 60°C

En base a estas condiciones originales, el diseño de Plackett Burman estudiará el efecto de desviaciones controladas a la metodología inicial. Para ello la nomenclatura del diseño experimental asigna a la condición inicial el valor 0, +1 al valor por sobre la condición original y -1 a aquella por debajo de ésta. De esta forma, los factores y niveles a estudiar son los siguientes:

Tabla 11. Tabla de factores y niveles estudio de robustez

Factor	-1	0	+1
A: Reactivo derivatizante	Sylon HTP		BSFTA
B: Tiempo de Reacción	30 min	40 min	50 min
C: Cantidad de Reactivo	80 uL	90 uL	100 uL
D: Temperatura de reacción	55°C	60°C	65°C

En el caso de la variable cualitativa *Reactivo derivatizante*, la asignación -1 y +1 es irrelevante.

La siguiente tabla muestra la matriz de diseño Plackett Burman para estudiar $= 4$ factores con 8 experimentos. Como un diseño Plackett Burman permite estudiar factores con $+ 1$ experimentos, esto quiere decir que con esta matriz se pueden evaluar $= 7$ factores. Como sólo deseamos estudiar 4, los otros 3 factores representarán variables *mudas* o *dummies* y nos permitirán estimar el error del diseño y la significancia de cada uno de los efectos.

Cada fila representa un experimento con las condiciones indicadas en la tabla anterior. Por ejemplo, el experimento 1 debe ser llevado a cabo utilizando BFSTA, 30 min de tiempo de reacción, 80 de reactivo y a 65°C.

Tabla 12. Matriz de diseño y respuesta analítica

Experimento	A	d1	B	d2	C	d3	D	R
1	+	-	-	+	-	+	+	100

2	+	+	-	-	+	-	+	90
3	+	+	+	-	-	+	-	100
4	-	+	+	+	-	-	+	90
5	+	-	+	+	+	-	-	80
6	-	+	-	+	+	+	-	70
7	-	-	+	-	+	+	+	70
8	-	-	-	-	-	-	-	70

El cálculo de cada efecto (E) es simple:

$$E = \frac{2[\sum(R+) - \sum(R-)]}{N}$$

donde $\sum(R+)$ representa la suma de las respuestas analíticas cuando el factor está en su nivel alto $(+)$, $\sum(R-)$ representa la suma de las respuestas analíticas cuando el factor está en su nivel bajo $(-)$ y N es el número de experimentos, en este caso $N = 8$. Como ejemplo, estimaremos el efecto del factor A:

$$E_A = \frac{2[100 + 90 + 100 + 80 - 90 - 70 - 70 - 70]}{8}$$

$$= \frac{70}{4}$$

$$= +17.5$$

El signo del efecto del factor A $E_A = +1.75$ indica que la respuesta analítica se incrementa cuando se mueve desde le nivel bajo (-) al nicel alto (+). La siguiente tabla muestra el efecto de cada uno de los factores excluyendo las variables mudas o *dummies*:

Tabla 13. Efectos de cada factor del estudio de robustez

Factor	Efecto
A: Reactivo derivatizante	17.5
d1	7.5
B: Tiempo de Reaccion	2.5
d2	2.5
C: Cantidad de Reactivo	-12.5
d3	2.5
D: Temperatura de reaccion	7.5

Sin embargo, si bien se obtiene un valor numérico de cada efecto, es necesario evaluar la significancia estadística de éstos a través de un análisis de varianza (ANOVA). Para cada factor la suma de los cuadrados es un diseño de dos niveles (+1 y -1) es dada por:

$$SS = \frac{N \times E^2}{4}$$

La suma de cuadrados para A, B, C y D son, por lo tanto; 6.125, 0.125, 3.125, and 1.125, respectivamente. Cada una de estas sumas de cuadrado sólo tiene un grado de libertad, de tal manera que sus varianzas (cuadrados medios) son equivalentes a las sumas de cuadrados. Análogamente, las SS para las variables *dummies* d1, d2 y D3 son 1.125, 0.125, and 0.125, respectivamente. La varianza de estas variables *dummies* es 0.458 la cual corresponde al promedio de las sumas de cuadrados. Por lo tanto, esta varianza es un estimador del error aleatorio del experimento. Ahora, cada uno de los factores individuales puede ser comparado con el error aleatorio mediante un test F estableciendo, arbitrariamente, un nivel de significancia de $\alpha = 0.05$. De esta forma, por ejemplo, para el factor A el $F^{A}_{calculado} = 6.125/0.458 = 13.37$. El valor crítico de F con 1 grado de libertad en el numerador y 3 en el denominador, con $\alpha = 0.05$ es $F_{critico} = 10.13$. Por lo tanto, ya que el $F^{A}_{calculado} > F_{critico}$, entonces, el efecto del factor A es estadísticamente significativo, por lo que este factor debe ser controlado rigurosamente durante el desarrollo de de un análisis.

La tabla siguiente entrega un resumen de los efectos y la significancia de cada factor:

Tabla 14. Tabla de efectos y significancia

Factor	Efecto	F calculado
A: Reactivo derivatizante	17.5	13.37
B: Tiempo de Reacción	2.5	0.27
C: Cantidad de Reactivo	-12.5	6.82
D: Temperatura de reacción	7.5	2.46

De la tabla se deduce que el único factor que tiene un efecto significativo es el factor A, es decir, el tipo de reactivo derivatizante. Como el signo del efecto del factor A es positivo y ya que la respuesta analítica estudiada es % de recuperación, entonces se deduce que se obtiene una mayor recuperación utilizando el nivel +1 de este factor, es decir, con el reactivo BSFTA. La siguiente figura describe la magnitud y signo de los efectos.

Algunas consideraciones de los estudios de robustez

68. Como la decisión sobre la significancia del efecto encontrado se basa en una prueba estadística, esta decisión estará afecta a todas las consideraciones de los tests estadísticos, por ejemplo, falsos positivos (error tipo I) y falsos negativos (error tipo II).

69. Para mejorar la estimación del error aleatorio en un diseño de Plackett-Burman, se sugiere siempre considerar al menos 3 variables *dummies*. Por ejemplo, si se quiere estudiar $n = 7$ factores, en vez de utilizar un diseño con 8 experimentos, se sugiere implementar un diseño de 12 experimentos (el cual permite estudiar $n - 1 = 11$ factores), de tal forma que con las $11 - 7 = 4$ variables *dummies* se estime el error aleatorio. La matriz de diseño para 12 experimentos se muestra en la siguiente tabla:

Tabla 15. Matriz de diseño Plackett-Burman de 12 experimentos para n = 7 factores

A	d1	B	d2	C	d3	D	d4	E	G	F
1	1	-1	1	1	1	-1	-1	-1	1	-1
-1	1	1	-1	1	1	1	-1	-1	-1	1
1	-1	1	1	-1	1	1	1	-1	-1	-1
-1	1	-1	1	1	-1	1	1	1	-1	-1
-1	-1	1	-1	1	1	-1	1	1	1	-1
-1	-1	-1	1	-1	1	1	-1	1	1	1
1	-1	-1	-1	1	-1	1	1	-1	1	1
1	1	-1	-1	-1	1	-1	1	1	-1	1
1	1	1	-1	-1	-1	1	-1	1	1	-1
-1	1	1	1	-1	-1	-1	1	-1	1	1
1	-1	1	1	1	-1	-1	-1	1	-1	1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

1. Especificidad/Selectividad

De acuerdo a la Eurachem/IUPAC la selectividad se define como:

El grado en el que un método puede ser utilizado para determinar analitos particulares en mezclas o matrices sin interferencias de otros componentes de comportamiento similar.

De acuerdo a esta definición y a las aclaraciones que han sido indicadas por otros organismos competentes (IUPAC), la selectividad no es equivalente a la especificidad. Esta última es definida como selectividad *absoluta*, es decir, el método analítico es específico para un mensurando cuando la señal instrumental es 100% selectiva para ese mensurando y, por lo tanto, libre de

todo tipo de interferencias. Métodos analíticos altamente selectivos hay muchos, sin embargo, son muy escasos los métodos específicos.

La discusión sobre la diferencia entre ambos términos es de larga data, sin embargo, se ha llegado al consenso de adoptar la definición de selectividad de un método analítico, tal como fue definida más arriba. La IUPAC, de hecho, no recomienda el uso del término especificidad. Más detalles sobre esta discusión pueden consultarse en la bibliografía adjunta.

Dicho esto, a pesar de las definiciones consensuadas, la guía de validación 2002/657/CE utiliza el término *especificidad*:

Capacidad de un método de distinguir entre el analito que se está midiendo y otras sustancias. Esta característica es ante todo una función de la técnica de medición descrita, pero puede variar en función del tipo de compuesto o de la matriz.

Ciertamente es una cuestión de nomenclatura, sin embargo, para ser consistentes con el vocabuario más moderno y de consenso, esta guía adoptará la definición de selectividad de IUPAC/Eurachem.

En general, las directrices de las distintas normas y guías proponen que el laboratorio investigue las potenciales sustancias interferentes previsibles, mediante el uso de compuestos homólogos o relacionados estructuralmente. Así también, indican que se debe estudiar aquellas interferencias que son producidas directamente por componentes de la matriz.

Cumplir completamente este objetivo es una tarea larga y muy costosa, por lo que el laboratorio debe diseñar, en base a conocimiento científico y/o experiencia, los estudios de selectividad que permitan concluir con cierto grado de confianza que la señal instrumental es lo suficientemente selectiva para la medición del analito.

A continuación, se entregan algunas pautas para diseñar estudios de selectividad de acuerdo a la técnica analítica:

70. Métodos cromatográficos sin detección espectrométrica:

- Riguroso análisis de reactivos (solventes, etc.), blancos reactivos (es decir, que han sido sometidos a todo proceso analítico) y blancos de matriz (en la medida que existan). En estos métodos cromatográficos, se recomienda obtener un cromatograma de cada uno de estos blancos, de tal manera visualizar potenciales interferencias de cada fuente.

- Se sugiere obtener mediante diferencia numérica, en el mismo software del cromatógrafo o en planilla de cálculo, un cromatograma neto, es decir, la diferencia

entre el blanco matriz y el blanco de reactivos. En el tiempo de retención del analito, este cromatograma no debiera presentar ninguna señal interferente (altura de peak menor a 3 veces la razón señal ruido).

- Si hay dudas debido a la presencia de picos solapados, se sugiere utilizar técnicas numéricas de tratamiento de señales cromatográficas para evaluar el grado de solapamiento (cromatograma derivado), con el fin de tomar acciones para mejorar la separación. Generalmente, estas técnicas numéricas están incorporadas en el software de tratamiento de datos del cromatógrafo.

7. Métodos cromatográficos con detección espectrométrica

- La confirmación de la identidad del analito debe seguir las directrices de la guía de validación se 2002/657/CE secciones 2.3.3 y 2.3.5.

- Utilice las herramientas del software de control cromatográfico y obtenga los parámetros de *pureza de peak* para cada analito en muestra matriz (AMDIS en espectrometría de masa, ángulo de pureza en DAD, razón isotópica en ICP-MS, etc.)

Para ambos tipos de técnicas cromatográficas se sugiere adicionar compuestos químicamente relacionados con el analito con el fin evaluar sus potenciales interferencias (isómeros, metabolitos, etc.)

8. Métodos confirmatorios para la determinación de metales

- La confirmación de la identidad del analito debe seguir las directrices de la guía de validación se 2002/657/CE secciones 2.3.3 y 2.3.5

Es posible estudiar la selectividad de un método analítico (cualquiera) en forma general través de la evaluación del efecto matriz. Este efecto se traduce en un aumento (o disminución) de la seña analítica en la matriz cuando se compara con el analito en solvente o solución ácida pura. Por ejemplo, un compuesta químico molecular y refractario al proceso e atomización

Si las interferencias de la matriz son proporcionales a la señal analítica, entonces, es un reflejo de un efecto proporcional o rotacional, por lo tanto, este efecto cambiará la pendiente de la función de calibración. Para evaluar este efecto se debe comparar la pendiente de la curva de calibración estándar (en solvente puro o solución ácida) con la pendiente de la curva de calibración obtenida por el método de adición conocida. Si las pendientes son iguales, entonces , el efecto de la matriz no es significativo. La siguiente figura describe el efecto matriz sobre la señal instrumental:

Si bien podría utilizarse los intervalos de confianza de ambas pendientes para evaluar, informalmente, la similitud entre ellas, la herramienta estadística idónea es análisis de covarianza (ANCOVA), la cual está disponible en la mayoría de los softwares estadísticos. Notar que el método de la adición conocida sólo es útil para efectos rotacional, no cuando existe un error sistemático.

2. Revalidación y/o verificación de un método analítico de acuerdo a esta guía.

Si el laboratorio de la red Sernapesca no ha validado un método analítico, es requerimiento que siga las directrices que entrega esta guía como un plan de validación mínimo.

Si un laboratorio tiene todas sus metodologías validadas debe verificar que ha validado todos los parámetros que se indican en esta guía, proveyendo evidencia estadística, metrológica y/o documental que cumple con el plan de validación mínimo descrito en este documento. Sernapesca comprende que existen otras aproximaciones a la validación de cada parámetro analítico, sin embargo, el laboratorio que ha utilizado estas aproximaciones debe evidenciar que ha logrado con ellas un nivel de entendimiento tal, que es equivalente al plan de validación propuesto en esta guía. A continuación, se indican algunos ejemplos de aproximaciones alternativas que son consistentes con esta guía:

71. Sesgo: Es posible evaluar el sesgo de un método participando en un ensayo colaborativo entre laboratorios, es decir, cuando todos los participantes utilizan exactamente el mismo método (ISO 5725-1). Sin embargo, para establecer el sesgo (metrológico) es necesario que la muestra utilizada posea un valor de referencia e incertidumbre, cuya asignación sea independiente de los participantes del ensayo colaborativo y, además, posea trazabilidad al Sistema Internacional de Unidades.

En el caso que el laboratorio participe en una ronda interlaboratorio o *proficiency test*, cuya asignación del valor de referencia es a través de la aplicación de métodos estadísticos de consenso (ISO 13528) y sin trazabilidad al SI, entonces el

sesgo calculado podría no reflejar la verdadera diferencia con el *valor verdadero* de concentración de la muestra. Precisamente, el valor de consenso es relativo a los participantes de la ronda y es un ejercicio de comparación interna, es decir, eventualmente el valor de consenso podría estar sesgado debido a que los laboratorios también lo están. Por lo tanto, si bien las rondas interlabotatorio son una excelente herramienta para evidencia el desempeño de un laboratorio, las conclusiones acerca del sesgo de un método obtenido a partir de valores de consenso tienen menor jerarquía que en aquellas rondas cuyo valor de referencia ha sido asignado en forma independiente a los participantes, con un método de alta jerarquía metrológica y trazabilidad al SI establecida.

Por ejemplo, la siguiente figura muestra la diferencia entre un valor de referencia obtenido por consenso entre los laboratorios (204 mg/kg Sn) y un valor asignado por el LGC (*Laboratory of Government Chemist*, el Instituto Nacional de Metrología del Reino Unido) de 247.8 ± 1.6 mg/kg Sn mediante dilución isotópica acoplada a espectrometría de masa (ID-ICPMS), un método primario relativo de la más alta jerarquía metrológica y con una completa cadena de trazabilidad al Sistema Internacional de Unidades. La diferencia del 17% es muy significativa desde el punto de vista metrológico, por lo tanto, lo será también la estimación del sesgo de una metodología dependiendo con cuál valor de referencia se compare.

Otra forma de evaluar sesgo es mediante la comparación del método a validar con un método cuyo sesgo es conocido (*Standard Method*) en un amplio rango de concentración. Se puede utilizar una regresión lineal entre los resultados analizados para varias muestras obtenidos por el método de referencia (X) y el método a validar (Y). En ausencia de sesgo se espera una pendiente $\beta_1 = 1$ e intercepto $\beta_0 = 0$. También es posible utilizar el método de Bland - Altman para llevar a cabo la comparación.

8. Precisión: Este parámetro (repetibilidad & reproducibilidad) puede ser obtenido también de la participación del laboratorio en un ensayo colaborativo. No es necesario que la muestra sea un material de referencia, basta que sea lo suficientemente homogénea y estable (ISO 5725-3).

9. Incertidumbre: Es posible obtener la incertidumbre de un método a partir de un diseño experimental como los describe la guía ISO 21748. Aunque no es un requerimiento de esta guía, se sugiere y alienta a los laboratorios a validar sus cálculos de incertidumbre mediante las directrices del Suplemento 1 de la ISO/GUM utilizando el Método de Monte Carlo, para lo cual existe software gratuito mantenido por centros internacionales de metrología en este link.

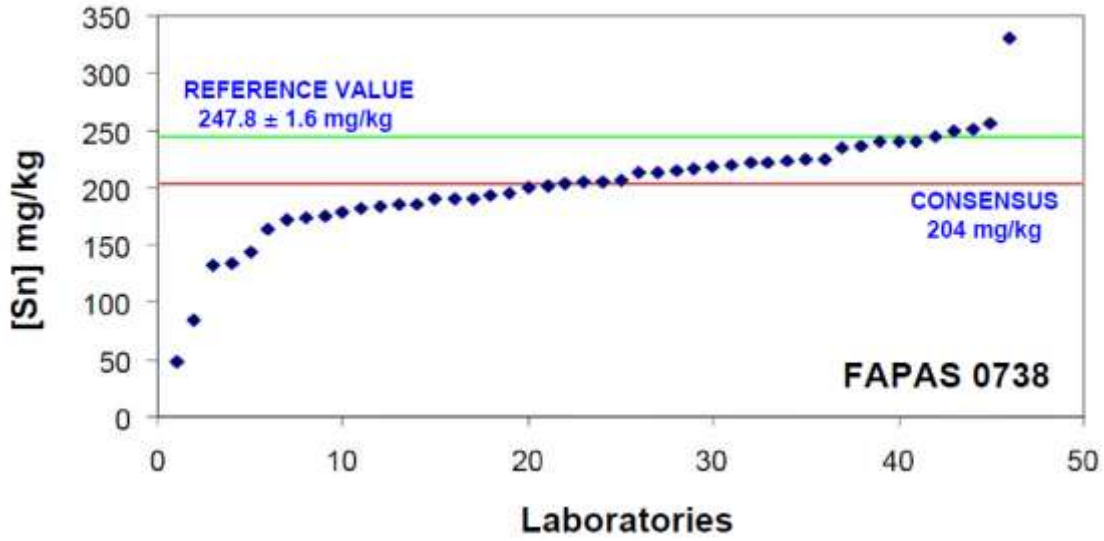

Figura 16. Diferencia entre el valor de consenso y el valor de referencia asignado por método primario IDMS. FAPAS 0738

Bibliografía

1. Albritton, D. L., A. L. Schmeltekopf, and R. N. Zare. 1976. "An Introduction to the Least-Squares Fitting of Spectroscopic Data." *ResearchGate* -1 (January): 1. https://www.researchgate.net/publication/234326696_An_Introduction_to_the_Least-Squares_Fitting_of_Spectroscopic_Data.

2. Alkemade, C. Th. J., W. Snelleman, G. D. Boutilier, and J. D. Winefordner. 1980. "A Review and Tutorial Discussion of Noise and Signal-to-Noise Ratios in Analytical Spectrometry—III. Multiplicative Noises." *Spectrochimica Acta Part B: Atomic Spectroscopy* 35 (5): 261–70. doi:10.1016/0584-8547(80)80089-9.

3. Alkemade, C. Th. J., W. Snelleman, G. D. Boutilier, B. D. Pollard, J. D. Winefordner, T. L. Chester, and N. Omenetto. 1978. "A Review and Tutorial Discussion of Noise and Signal-to-Noise Ratios in Analytical Spectrometry—I. Fundamental Principles of Signal-to-Noise Ratios." *Spectrochimica Acta Part B: Atomic Spectroscopy* 33 (8): 383–99. doi:10.1016/0584-8547(78)80049-4.

4. Andersen, Jens E. T. 2017. "The Standard Addition Method Revisited." *TrAC Trends in Analytical Chemistry* 89 (Supplement C): 21–33. doi:10.1016/j.trac.2016.12.013.

5. Andrade, J. M., and M. P. Gómez-Carracedo. 2013. "Notes on the Use of Mandel's Test to Check for Nonlinearity in Laboratory Calibrations." *Analytical Methods* 5 (5): 1145–9. doi:10.1039/C2AY26400E.

6. Andrade, J. M., J. Terán-Baamonde, R. M. Soto-Ferreiro, and A. Carlosena. 2013. "Interpolation in the Standard Additions Method." *Analytica Chimica Acta* 780 (May): 13–19. doi:10.1016/j.aca.2013.04.015.

7. Barwick, Vicki J, Stephen L. R Ellison, and Ben Fairman. 1999. "Estimation of Uncertainties in ICP-MS Analysis: A Practical Methodology." *Analytica Chimica Acta* 394 (2): 281–91. doi:10.1016/S0003-2670(99)00302-5.

8. Barwick, Vicki, Stephen L. R. Ellison, Elin Gjengedal, Bertil Magnusson, Olivier Molinier, Marina Patriarca, Lorens Sibbesen, Nicole Vanlaethem, and Isabelle Vercruysse. 2017. "Method Validation in Analytical Sciences: Discussions on Current Practice and Future Challenges." *Accreditation and Quality Assurance*, September, 1–11. doi:10.1007/s00769-017-1286-4.

9. Bland, J. M., and D. G. Altman. 1995. "Comparing Methods of Measurement: Why Plotting Difference Against Standard Method Is Misleading." *The Lancet* 346 (8982): 1085–7. doi:10.5555/uri:pii:S0140673695917489.

10. Boque, R., F. X. Rius, and D. L. Massart. 1994. "Straight Line Calibration: Something More Than Slopes, Intercepts, and Correlation Coefficients." *Journal of Chemical Education* 71 (3): 230. doi:10.1021/ed071p230.

11. Boutilier, G. D., B. D. Pollard, J. D. Winefordner, T. L. Chester, and N. Omenetto. 1978. "A Review and Tutorial Discussion of Noise and Signal-to-Noise Ratios in Analytical Spectrometry—II. Fundamental Principles of Signal-to-Noise Ratios." *Spectrochimica Acta Part B: Atomic Spectroscopy* 33 (8): 401–16. doi:10.1016/0584-8547(78)80050-0.

12. Cowen, Simon, and Stephen L. R. Ellison. 2006. "Reporting Measurement Uncertainty and Coverage Intervals Near Natural Limits." *Analyst* 131 (6): 710–17. doi:10.1039/B518084H.

13. Cuadros-Rodríguez, Luis, Laura Gámiz-Gracia, Eva M Almansa-López, and Juan M. Bosque-Sendra. 2001a. "Calibration in Chemical Measurement Processes. II. A Methodological Approach." *TrAC Trends in Analytical Chemistry* 20 (11): 620–36. doi:10.1016/S0165-9936(01)00111-X.

14. Cuadros-Rodríguez, Luis, Laura Gámiz-Gracia, Eva Almansa-López, and Jesús Laso-Sánchez. 2001b. "Calibration in Chemical Measurement Processes: I. A Metrological Approach." *TrAC Trends in Analytical Chemistry* 20 (4): 195–206. doi:10.1016/S0165-9936(00)00093-5.

15. Currie, Lloyd A. 1999. "Detection and Quantification Limits: Origins and Historical Overview1." *Analytica Chimica Acta* 391 (2): 127–34. doi:10.1016/S0003-2670(99)00105-1.

16. Danzer, Klaus, and L. A. Currie. 2009. "Guidelines for Calibration in Analytical Chemistry. Part I. Fundamentals and Single Component Calibration (IUPAC Recommendations 1998)." *Pure and Applied Chemistry* 70 (4): 993–1014. doi:10.1351/pac199870040993.

17. Ellison, Stephen L. R. 1998. "Using Validation Data for ISO Measurement Uncertainty estimationPart 1. Principles of an Approach Using Cause and Effect Analysis." *Analyst* 123 (6): 1387–92. doi:10.1039/A706946D.

18. Ellison, S.L.R. 2006 "In defence of the correlation coefficient" Accred Qual Assur 11: 146. https://doi.org/10.1007/s00769-006-0087-y

19. ———. 2014. "Implementing Measurement Uncertainty for Analytical Chemistry: The Eurachem Guide for Measurement Uncertainty." *Metrologia* 51 (4): S199. doi:10.1088/0026-1394/51/4/S199.

20. Ellison, Stephen L. R., and Michael Thompson. 2008. "Standard Additions: Myth and Reality." *Analyst* 133 (8): 992–97. doi:10.1039/B717660K.

21. Ellison, Stephen L. R., Vicki J. Barwick, and Trevor J. Duguid Farrant. 2009. *Practical Statistics for the Analytical Scientist*. http://pubs.rsc.org/en/content/ebook/9780854041312.

22. Galbán, Javier, and Carlos Ubide. 2007. "Uncertainty Due to the Quantification Step in Analytical Methods." *Talanta* 71 (3): 1339–44. doi:10.1016/j.talanta.2006.07.031.

23. Hartmann, C., J. Smeyers-Verbeke, W. Penninckx, and D.L. Massart. 1997. "Detection of Bias in Method Comparison by Regression Analysis." *Analytica Chimica Acta* 338 (1–2): 19–40. doi:10.1016/S0003-2670(96)00341-8.

24. Hayash, Yuzuru, and Rieko Matsuda. 1994. "Uncertainty Structure of Dilution with Pipette and Volumetric Flask." *Analytical Sciences* 10 (6): 881–88. doi:10.2116/analsci.10.881.

25. Hayashi, Y., and R. Matsuda. 1995. "Prediction of Precision from Signal and Noise Measurement in Liquid Chromatography: Mathematical Relationship Between Integration Domain and Precision." *Chromatographia* 41 (1-2): 75–83. doi:10.1007/BF02274198.

26. Hayashi, Yuzuru, and Nien Fan Zhang. 2015. "Evaluation of Measurement Precision from Stationary Baseline Noise in Instrumental Analyses." *Analytical Sciences* 31 (12): 1219–24. doi:10.2116/analsci.31.1219.

27. "ISO 13528:2015 - Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparison." 2017. Accessed October 17. https://www.iso.org/standard/56125.html.

28. "ISO Guide 35:2017 - Reference Materials – Guidance for Characterization and Assessment of Homogeneity and Stability." 2017. Accessed October 17. https://www.iso.org/standard/60281.html.

29. "ISO/IEC Guide 98-3:2008 - Uncertainty of Measurement – Part 3: Guide to the Expression of Uncertainty in Measurement (GUM:1995)." 2017. Accessed October 17. https://www.iso.org/standard/50461.html.

30. Joglekar, G., J. H. Schuenemeyer, and V. Lariccia. 1989. "Lack-of-Fit Testing When Replicates Are Not Available." *The American Statistician* 43 (3): 135–43. doi:10.1080/00031305.1989.10475641.

31. Kadis, Rouvim. 2017. "Evaluation of the Measurement Uncertainty: Some Common Mistakes with a Focus on the Uncertainty from Linear Calibration." *Journal of Chromatography A* 1499 (May): 226–29. doi:10.1016/j.chroma.2017.03.078.

32. Kallner, Anders. 2014. "Comprehensive Method Comparisons: Getting More from the Data." *Accreditation and Quality Assurance* 19 (6): 451–57. doi:10.1007/s00769-014-1089-9.

33. Kruve, Anneli, Riin Rebane, Karin Kipper, Maarja-Liisa Oldekop, Hanno Evard, Koit Herodes, Pekka Ravio, and Ivo Leito. 2015a. "Tutorial Review on Validation of Liquid Chromatography–mass Spectrometry Methods: Part I." *Analytica Chimica Acta* 870: 29–44. doi:10.1016/j.aca.2015.02.017.

34. ———. 2015b. "Tutorial Review on Validation of Liquid Chromatography–mass Spectrometry Methods: Part II." *Analytica Chimica Acta* 870: 8–28. doi:10.1016/j.aca.2015.02.016.

35. Lambert, Diane, Bruce Peterson, and Irma Terpenning. 1991. "Nondetects, Detection Limits, and the Probability of Detection." *Journal of the American Statistical Association* 86 (414): 266–77. doi:10.1080/01621459.1991.10475030.

36. Larsen, Ingvar L., Norbert A. Hartmann, and Jerome J. Wagner. 1973. "Estimating Precision for the Method of Standard Additions." *Analytical Chemistry* 45 (8): 1511–3. doi:10.1021/ac60330a005.

37. Leardi, Riccardo. 2009. "Experimental Design in Chemistry: A Tutorial." *Analytica Chimica Acta*, Fundamental and Applied Analytical Science. A Special Issue In Honour of Alan Townshend.FAAS 2009, 652 (1–2): 161–72. doi:10.1016/j.aca.2009.06.015.

38. Liddell, Peter R. 1976. "Noise at Detection Limit Levels in Atomic Absorption Flame Spectrometry." *Analytical Chemistry* 48 (13): 1931–3. doi:10.1021/ac50007a031.

39. Lindsay, W. S., and and R. F. Picone. 1984. "Components of Variance Analysis: A Useful Statistical Technique For Analytical Chemistry." *Analytical Letters* 17 (15): 1731–42. doi:10.1080/00032718408065346.

40. Long, Gary L., and J. D. Winefordner. 1983. "Limit of Detection. A Closer Look at the IUPAC Definition." *Analytical Chemistry* 55 (7): 712A–724A. doi:10.1021/ac00258a001.

41. Magnusson, Bertil, and Stephen L. R. Ellison. 2008. "Treatment of Uncorrected Measurement Bias in Uncertainty Estimation for Chemical Measurements." *Analytical and Bioanalytical Chemistry* 390 (1): 201–13. doi:10.1007/s00216-007-1693-1.

42. Matsuda, R. 1998. "Deductive Prediction of Precision in Measurement, Calibration, and Standard Addition Method in Atomic Absorption Spectrometry for Cadmium." *Analytical Chemistry* 70 (2): 319. doi:10.1021/ac970616f.

43. Melucci, Dora. 2008. "Manual Data Processing in Analytical Chemistry: Linear Calibration." *Journal of Chemical Education* 85 (10): 1346. doi:10.1021/ed085p1346.

44. "Method Validation." 2017. Accessed October 17. https://www.eurachem.org/index.php/publications/guides/mv.

45. Miller, James, and Jane C. Miller. 2010. *Statistics and Chemometrics for Analytical Chemistry*. 6 edition. Harlow: Pearson Education Canada.

46. Phillips, John B. 1982. "Signal Processing Techniques in Analytical Instruments." *TrAC Trends in Analytical Chemistry* 1 (7): 163–66. doi:10.1016/0165-9936(82)80051-4.

47. Prenesti, Enrico, and Fabio Gosmaro. 2015. "Trueness, Precision and Accuracy: A Critical Overview of the Concepts as Well as Proposals for Revision." *Accreditation and Quality Assurance* 20 (1): 33–40. doi:10.1007/s00769-014-1093-0.

48. "Quality in Analytical Chemistry." 2017. Accessed October 17. https://www.eurachem.org/index.php/publications/guides/qa.

49. "Quantifying Uncertainty." 2017. Accessed October 17. https://www.eurachem.org/index.php/publications/guides/quam.

50. Raposo, Francisco. 2016. "Evaluation of Analytical Calibration Based on Least-Squares Linear Regression for Instrumental Techniques: A Tutorial Review." *TrAC Trends in Analytical Chemistry* 77 (March): 167–85. doi:10.1016/j.trac.2015.12.006.

51. Richter, W. 1997. "Primary Methods of Measurement in Chemical Analysis." *Accreditation and Quality Assurance* 2 (8): 354–59. doi:10.1007/s007690050165.

52. Riu, Jordi, and F. Xavier Rius. 1996. "Assessing the Accuracy of Analytical Methods Using Linear Regression with Errors in Both Axes." *Analytical Chemistry* 68 (11): 1851–7. doi:10.1021/ac951217s.

53. Sayago, Ana, Maravillas Boccio, and Agustin Asuero. 2004. "Fitting Straight Lines with Replicated Observations by Linear Regression: The Least Squares Postulates." *Critical Reviews in Analytical Chemistry* 34 (1): 39–50. doi:10.1080/10408340490273744.

54. Schwartz, Lowell M. 1976. "Nonlinear Calibration Curves." *Analytical Chemistry* 48 (14): 2287–9. doi:10.1021/ac50008a064.

55. Silva, Ricardo J. N. Bettencourt da, and M. Filomena Camões. 2010. "The Quality of Standards in Least Squares Calibrations." *Analytical Letters* 43 (7-8): 1257–66. doi:10.1080/00032710903518674.

56. Stephens, T. W. 1979. "Necessity of a Priori Design Considerations in Experimental Statistical Determination of Deviation from Linearity." *Clinical Chemistry* 25 (12): 2052–5. http://clinchem.aaccjnls.org/content/25/12/2052.

57. "Terminology in Analytical Measurement." 2017. Accessed October 17. https://www.eurachem.org/index.php/publications/guides/terminology-in-analytical-measurement.

58. Thompson, Michael. 2012. "Precision in Chemical Analysis: A Critical Survey of Uses and Abuses" 4 (6): 1598–1611. doi:10.1039/C2AY25083G.

59. ———. 2016. "The Comparison Between Reproducibility Standard Deviations from Collaborative Trials and Proficiency Tests: A Preliminary Study from Food Analysis." *Analytical Methods* 8 (4): 742–46. doi:10.1039/C5AY02890F.

60. Thompson, Michael, and Stephen L. R. Ellison. 2013. "Towards an Uncertainty Paradigm of Detection Capability." *Analytical Methods* 5 (21): 5857–61. doi:10.1039/C3AY41209A.

61. Thompson, Michael, and Richard J. Howarth. 1978. "A New Approach to the Estimation of Analytical Precision." *Journal of Geochemical Exploration* 9 (1): 23–30. doi:10.1016/0375-6742(78)90035-3.

62. Thompson, Michael, Steven L R Ellison, Aleš Fajgelj, Paul Willetts, and Roger Wood. 2009. "Harmonized Guidelines for the Use of Recovery Information in Analytical Measurement." *Pure and Applied Chemistry* 71 (2): 337–48. doi:10.1351/pac199971020337.

63. Utts, Jessica M. 1982. "The Rainbow Test for Lack of Fit in Regression." *Communications in Statistics - Theory and Methods* 11 (24): 2801–15. doi:10.1080/03610928208828423.

64. Valcárcel, Miguel, and Angel Ríos. 1999. "A Metrological Hierarchy for Analytical Chemistry." *TrAC Trends in Analytical Chemistry* 18 (1): 68–75. doi:10.1016/S0165-9936(98)00088-0.

65. Wang, Daniel, and Michael Conerly. 2003. "Evaluation of Three Lack of Fit Tests in Linear Regression Models." *Journal of Applied Statistics* 30 (6): 683–96. doi:10.1080/0266476032000053763.

66. Wong, Siu Kay. 2016. "Review of the New Edition of ISO 13528." *Accreditation and Quality Assurance*, June, 1–6. doi:10.1007/s00769-016-1217-9.

67. Zwanziger, Heinz W., and Costel Sârbu. 1998. "Validation of Analytical Methods Using a Regression Procedure." *Analytical Chemistry* 70 (7): 1277–80. doi:10.1021/ac970926y.

Checklist Mínimo

Parámetro	Verificación
Alcance	1. Definición clara e inequívoca del alcance del método en el informe de validación: "Determinación de cloranfenicol en salmón fresco por ASE-NCI-GC/MS para un rango de concentración entre 0.3 ug/kg y 1.0 ug/kg" (en el alcance, no en el título del informe) 2. Notar que el rango de trabajo asume una lectura directa de la muestra, es decir, sin dilución. 3. Los alcances genéricos no están permitidos: "Determinación de metales en choritos por AAS" ¿Cuáles metales? ¿producto fresco o congelado?¿Cuál técnica espectrofotométrica: FAAS, ETAAS, etc.?
Linealidad	1. Al menos n = 6 calibrantes para el test de Mandel o n = 5 calibrantes en duplicado para el test de carencia de ajuste. Los calibrantes deben ser equidistantes 2. Verificar que el rango lineal validado satisfaga todos los requerimientos de MRL de los distintos mercados de destino. De lo contrario, salvo que se diluya la muestra antes de su lectura, el laboratorio debe establecer y validar el número de rangos que sean necesarios. 3. Definición del tipo de calibración:

	estándar, en matriz, uso de estándar interno, etc. 4. Gráfico de la curva de calibración
Precisión	1. Repetibilidad y precisión intermedia calculada a partir de ISO 5725 o a partir de datos de QAQC (sólo para repetibilidad). 2. Si se utiliza ISO 5725 definir cuál ha sido el factor estudiado (analista, instrumento, etc.) 3. (Optativo): cálculo de r: la máxima diferencia tolerable entre duplicados de análisis.
Límite de detección	1. LoD calculado de acuerdo a esta guía. Al menos n = 10 muestras blanco para la estimación
Límite de cuantificación	1. LoQ calculado de acuerdo a esta guía. Al menos n = 10 muestras blanco para la estimación
Selectividad	1. Riguroso análisis de reactivos (solventes, etc.), blancos reactivos (es decir, que han sido sometidos a todo proceso analítico) y blancos de matriz (en la medida que existan). 2. En métodos acoplados a MS evaluar las posibles interferencias isobáricas (misma masa) 3. (Optativo-Ideal): Comparación entre método de calibración estándar (en

	solvente o matriz ácida) con método de adición conocida.
Incertidumbre	1. Evaluación de las principales fuentes de incertidumbre mediante alguna de las guías ISO propuestas en esta guía. Idealmente, guía ISO-GUM. 2. Debe presentar un *budget* de incertidumbre, es decir, una tabla donde se indique la fuente y la incertidumbre estándar asociada. 3. Finalmente, se debe presentar la incertidumbre expandida con un 95% de confianza (k = 2).
Sesgo/Recuperación	4. Evaluación de un material de referencia certificado. Aplicación del test de sesgo propuesto en la guía 5. Evaluación de recuperación
Robustez	1. Identificación de los principales factores, cuali y/o cuantitativos, que podrían tener una influencia en el desempeño de la metodología 2. Una vez identificados estos factores diseñar un experimento de acuerdo a las directrices de esta guía, con el fin de evaluar la significancia de los efectos. Esta guía entrega una pauta, sin embargo, es posible hacer el mismo estudio mediante otras aproximaciones igualmente válidas, las cuales deben estar completamente

	documentadas.
Informe de validación	1. Debe contener cada uno de los puntos/parámetros evaluados 2. El laboratorio debería entregar evidencia de cromatogramas, espectros de masa, etc. que den soporte a los resultados de la validación. 3. Si el laboratorio ha utilizado otra norma/documento para abordar alguno de los parámetros de validación, deberá entregar una copia indicando cuál ha sido la aproximación utilizada. 4. (Optativo): El laboratorio debe indicar/entregar las planillas de cálculo y/o resultados de análisis estadísticos realizados durante la validación.