

UNA EXCESIVA CONFIANZA EN EL COEFICIENTE DE CORRELACIÓN DE PEARSON EN LAS INTERPOLACIONES

Gabino Estevez-Delgado*, Joaquín Estevez-Delgado, Julissa Rodríguez-Coss,
Itzia Alejandra Bonilla-Paz.

Universidad Michoacana de San Nicolás de Hidalgo.

[*gabino.estevez@umich.mx](mailto:gabino.estevez@umich.mx)

Resumen

La robustez de los procesos que se atienden en los laboratorios clínicos, de alimentos o de ensayos de calibración es una de las condiciones iniciales que debe atenderse, de ello depende el poder llevar a cabo mediciones sujetas a las normativas o referencias internacionales tales como las ISO 15189, ISO 22000, ISO 17025, entre otras que incluyen el cálculo de la incertidumbre, tal como lo demanda la guía del cálculo de la incertidumbre, GUM. Su intención es garantizar que los procesos de la incertidumbre son trazables al sistema internacional de unidades, pero además es importante garantizar la reproducibilidad de los fenómenos. La parte crucial del cálculo de la incertidumbre está ligado a la ley de propagación de errores cuyo elemento central tiene su fundamento en la generación de una ley geométrica que nos permite analizar cómo cada variable incrementa la variabilidad del proceso de manera geométrica, Una de las consideraciones que tenemos cuando obtenemos la incertidumbre expandida o variabilidad final del proceso es la obtención de valores interpolados sobre valores que no son los reportados en los materiales de referencia o valores de calibradores, en el caso biológico, lo que demanda el uso de interpolaciones, que se evalúan a través de valores de la mejor aproximación dada por el coeficiente de correlación de Pearson. En esta investigación se encuentra que la interpolación realizada en los límites de detección la interpolación no siempre es la dada por el valor más cercano a la unidad, sino que está sujeta a la parte

geométrica de la ley de propagación de errores para la variabilidad, siendo entonces importante la homocedasticidad de los datos.

Palabras clave: Interpolación polinomial y logarítmica, coeficiente de correlación de Pearson, Ley de propagación de errores, incertidumbre y error de medición

Introducción

Los modelos probabilísticos sobre los que se apoyan algunas disciplinas de aplicación se ha hecho cada vez más extensivos, lo que conlleva a regular los procesos de evaluación de la conformidad, la generación de regulaciones que involucran la toma de decisiones. El uso de áreas de las matemáticas que son de carácter probabilístico inicia con la propia determinación del tamaño de muestra, lo que garantiza la robustez de las mediciones y en consecuencia la confianza en lo que se realiza (Acree, 2021).

Una de las ciencias de aplicación es la ciencia de la medida llamada metrología, en la que el análisis de la variabilidad es un aspecto sumamente importante, considerado como parte de las regulaciones a nivel internacional. La generación de conceptos como la trazabilidad da lugar a otros que están íntimamente ligados a la estadística paramétrica, debido a que se busca por un lado homogeneidad de los datos y por otro lado isotropía en las mediciones a lo largo de periodos de tiempo determinados (Wei, *et al.*, 2021).

A la par de conceptos como los requeridos por la metrología científica, del campo probabilístico, se requiere en ocasiones otro tipo de metodologías, tales como lo es el uso de la interpolación. Siendo de interés particular las de las llevadas a cabo de manera periódica como es el caso de las valoraciones que se llevan dentro del campo clínico o químico con las llamadas curvas de calibración (Budhram *et al.*, 2022). El uso de estas permite el verificar algún resultado del que se tiene dudas,

pero a la par permite garantizar que un instrumento emite resultados técnicamente válidos.

La estadística paramétrica garantiza que debido a la robustez de las muestras las interpolaciones realizadas son válidas mediante la evaluación de la homocedasticidad de las mediciones descritas por coeficientes como es el caso del coeficiente de interpolación de Pearson, por lo que de manera rutinaria se usa desde hace tiempo con tal familiaridad que nadie cuestiona su uso si da cumplimiento al tamaño mínimo de muestra a través de pruebas de hipótesis que garantizan la normalidad tales como las pruebas de Shapiro Wilks o la de Kolmogorov Smirnov. Sin embargo, la determinación de curvas de manera general con intervalos extensos de datos no necesariamente va a conducir a la mejor evaluación, pero que además poco se cuestiona (Wei *et al.*, 2021).

En el caso particular de la magnitud de óptica, como se observará, se identifica que se puede generar aproximaciones a un conjunto de datos en el que los errores crecen a medida que se toman intervalos de datos que responden a leyes particulares, tales como la ley de Lambert Beer.

Objetivo

Establecer las condiciones iniciales para poder determinar el tamaño mínimo de muestra cuando la heterocedasticidad se presenta, como es el caso de los materiales transmitancia, que además pudieran generar limitantes en su concentración, lo que pudiera generar una incorrecta estimación en las interpolaciones, pese a un coeficiente de correlación de Pearson cercano, en valor absoluto a la unidad.

Materiales y métodos

Como parte del análisis alrededor de la presente investigación tomaremos un material de absorbanza óptica neutra a 90% de transmitancia, cuyo material es trazable al Sistema Internacional de Unidades. Sin embargo, para fines de confidencialidad mantiene un corrimiento en los valores presentados, con lo que se mantiene su comportamiento, pero permite la confidencialidad de la información, además de dar cumplimiento a la aplicación de la referencia ISO 17025, en cuanto a la exigencia de la imparcialidad (ISO 17025, 2017).

Tabla 1. Material certificado de densidad óptica neutra en transmitancia.

Longitud de onda	MR 90
225	91.4624
250	92.1254
280	92.6144
340	93.1844
360	93.2714
400	93.4324
405	93.4694
440	93.5814
450	93.6024
465	93.6444
490	93.6944
546.1	93.7964
550	93.7954
590	93.8444
620	93.8804
630	93.9014
635	93.9014
650	93.9044
690	93.9614
700	93.9594
750	93.9964

Al realizar la interpolación con la información contenida arrojada por el espectrofotómetro, luego de realizar las mediciones, se obtuvo el siguiente gráfico (Figura 1).

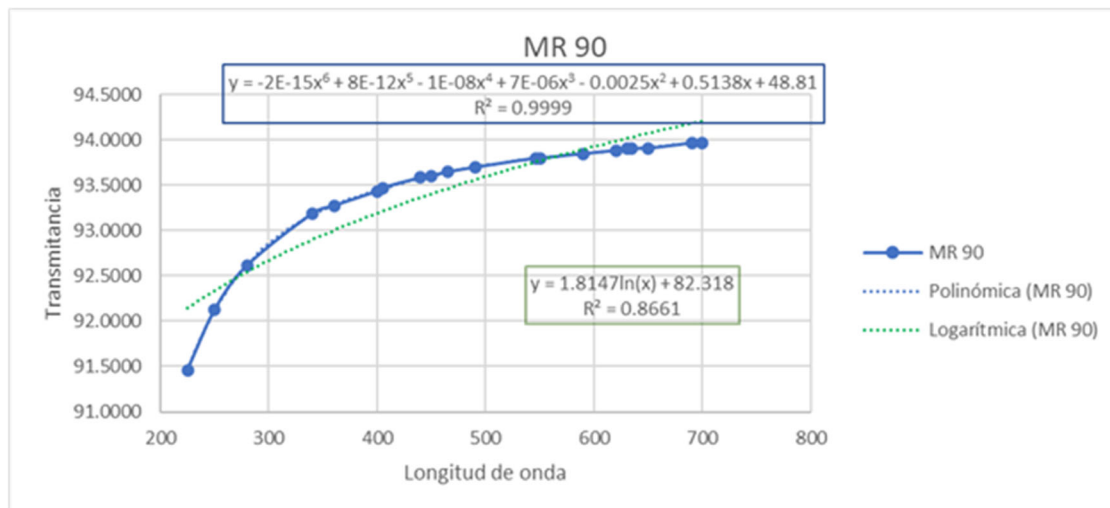


Figura 1. Gráfica de barrido para el material de 90% de transmitancia con interpolaciones

Resultados

Como se observa la interpolación más apropiada es la polinomial, dado que el coeficiente de correlación de determinación es mejor que el considerado para el caso logarítmico.

Es innegable que el proceso de interpolación no deja lugar a dudas sobre la elección, particularmente por el proceso de calibración, mantiene la robustez en los instrumentos ópticos, dada la repetibilidad y la velocidad de barrido descrita.

Es en este momento en el que la presente investigación indagara en aspectos finos sobre la interpolación y la robustes adecuada, que sin duda configura aspectos adicionales a la ley de propagación de errores, ¿Qué ocurre si la interpolación se demanda alrededor de 230 mn, mismo que no se tiene directamente descrito en el valor medido?

Realizando las aproximaciones alrededor del punto solicitado, se presentan los siguientes resultados en la Tabla 2.

Tabla 2. Resultado de interpolación alrededor de 230 nm.

Longitud de onda	MR 90	Aproximación polinomial	Aproximación logarítmica	Error para aproximación polinomial	Error para aproximación logarítmica
225	91.4624	91.3813	92.1466	0.0811	0.6842
250	92.1254	90.3459	92.3378	1.7794	0.2124
280	92.6144	87.1514	92.5435	5.4630	0.0709

Como se puede ver el error alrededor del valor, mediante el método polinomial crece el error en dos valores, lo que conlleva considerar que hay valores geométricos sobre los que debemos de reflexionar, sobre el tamaño de muestra y las aproximaciones polinomiales con respecto a otras, como el caso de las aproximaciones logarítmicas e inclusive sobre la propia relación con valores de relaciones entre conceptos ópticos como la relación de la transmitancia y la absorbancia.

Discusión

Como es bien conocido la robustes es un aspecto que impacta directamente en los parámetros tales como la linealidad o el propio ámbito de trabajo. Como se ha sido observado desde hace más de una década, Estevez (2012) y otros

investigadores han observado que se requiere que las interpolaciones pueden tener aspectos negativos cuando una mala planeación en el número de datos no se da de manera adecuada, particularmente, se ha observado la necesidad de retomar los valores en zonas en las que la homocedasticidad empieza a perderse, condiciones que se reafirman en esta investigación y que generan alertas en la evaluación de la interpolación.

Conclusión

Determinar el tamaño de muestra adecuada en cualquier investigación no es una tarea fácil, pese a que existen diferentes formalismos estadísticos, inclusive se ha generado investigaciones sobre el comportamiento de las muestras a fin de diferenciar el área de aplicación estadística, pudiendo ser no paramétrica cuando la cantidad de información no es la adecuada, y paramétrica en caso aceptable. Los trabajos pioneros realizados por Kolmogorov dieron lugar a realizar investigaciones en el contexto de la robustez de las muestras, desde luego que antecede otros trabajos como los de Gauss, D' Moivre, Laplace, Bernoulli, entre los más destacados.

Los trabajos realizados por Shapiro y Wilks para determinar la normalidad de una muestra han dado un impulso a determinación del análisis de trabajos en investigación, particularmente la fusión sus trabajos abonaron a la teoría de las pruebas de hipótesis sobre la normalidad.

Podríamos hacer un recorrido sobre las pruebas de hipótesis que abonan a la normalidad, sin embargo, como hemos podido constar de los resultados presentados, que además permiten verificar el cumplimiento de la hipótesis planteada; en esta investigación el tamaño de muestra en los materiales de referencia es ceñido a otros formalismos ligados al comportamiento de la materia.

Se concluye, que en efecto el material de óxido de didimio requiere ser analizado no solo por el comportamiento del tamaño mínimo de muestra sino que además requiere ser analizado a partir de aspectos como la homocedasticidad, dado que como vimos en los resultados presentados al momento de interpolar en áreas específicas del espectro electromagnético el comportamiento no se sujeta a consideraciones de interpolación descritas para el coeficiente de correlación de Pearson.

Se observó de manera puntual, que la interpolación global del material de 90% de transmitancia mantenía una mejor aproximación polinomial contrastando con la aproximación logarítmica, por el coeficiente de determinación, pero finalmente se distinguió que era insuficiente, dado que los errores para las aproximaciones crecen más con el mejor estimado para la determinación de Pearson, resultando conveniente ser cuidadosos con materiales que tienen regiones lineales y que posiblemente en áreas como la calculada es de carácter lineal en el intervalo de trabajo demandado.

Agradecimientos

Agradecemos el apoyo brindado por el laboratorio de Biofísica del Instituto de Física y Matemáticas de la Universidad Michoacana de San Nicolás de Hidalgo. Así como la Coordinación de Investigación Científica por los apoyos otorgados para llevar a cabo la investigación.

Referencias

Acree, M. C. (2021). *In The Myth of Statistical Inference*. En M. C. Acree, *Nineteenth-Century Developments in Statistics*. (págs. 131-158). Springer, Cham.

Recuperado el 18 de septiembre de 2021, de https://link.springer.com/chapter/10.1007/978-3-030-73257-8_5

Adatrao, S., Sciacchitano, A., van der Velden, S., van der Meulen, M. J., & Bordes, M. C. (2021). *Design of experiments: a statistical tool for PIV uncertainty quantification*. En U. Quantification (Ed.), In 14th International Symposium on Particle Image Velocimetry. 1. Illinois Tech. doi: <https://doi.org/10.18409/ispiv.v1i1.69>

Adeniran, A.T., Faweya, O., Ogunlade, T.O., & Balogun, K.O. (2020). Derivation of Gaussian Probability Distribution: A New Approach. *Applied Mathematics*, 11(06):436.

Budhram, A., Yang, L., Bhayana, V., Mills, J.R., & Dubey, D. (2022). Clinical sensitivity, specificity, and predictive value of neural antibody testing for autoimmune encephalitis. *The Journal of Applied Laboratory Medicine*, 71(1): 350-356.

Estevez, D.G., & Estvez, D.J. (2012). *Explorando el lado oscuro en las técnicas de interpolación*. XIII Evento Internacional MATECOMPU'2012, "La enseñanza de la matemática, la estadística y la computación", III Congreso Internacional "AL. Habana, Cuba.

ISO 17025. (2017). *ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratorio*. (3rd ed.). Geneve: ISO. Recuperado el 09 de septiembre de 2021

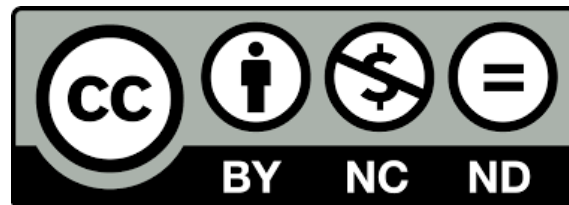
Wei, P.L., Valdebenito, M., & Beer, M. (2021). Bayesian probabilistic propagation of imprecise probabilities with large epistemic uncertainty. *Mechanical Systems and Signal Processing*, 149: 107219.



D. R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO.

Excepto donde se indique lo contrario esta obra está bajo una licencia Creative Commons Atribución No comercial, No derivada, 4.0 Internacional (CC BY NC ND 4.0 INTERNACIONAL).

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



ENTIDAD EDITORA

Facultad de Estudios Superiores Cuautitlán.

Av. Universidad 3000, Universidad Nacional Autónoma de México, C.U., Delegación Coyoacán, C.P. 04510, Ciudad de México.

FORMA SUGERIDA DE CITAR:

Estevez-Delgado, G., Estevez-Delgado, J., Rodríguez-Coss, J., y Bonilla-Paz, I. A. (2022). Una excesiva confianza en el coeficiente de correlación de pearson en las interpolaciones. *MEMORIAS DEL CONGRESO NACIONAL DE TECNOLOGÍA (CONATEC)*, Año 5, No. 5, septiembre 2022 - agosto 2023. Facultad de Estudios Superiores Cuautitlán. UNAM.

https://tecnicosacademicos.cuautitlan.unam.mx/CongresoTA/memorias2022/mem2022_ExtensioPaper7.html