
APLICATIVO WEB 'validaR' PARA EL TRATAMIENTO ESTADÍSTICO DE DATOS GENERADOS EN LABORATORIOS DE ANÁLISIS QUÍMICO

Cristhian Paredes¹, Diego Ahumada²

¹Grupo de Investigación en Metrología Química y Bioanálisis - Instituto Nacional de Metrología de Colombia
Avenida Carrera 50 No. 26-55, Interior 2, Bogotá D.C - Colombia
Tel: (1) 245 2222, correo: caparedes@inm.gov.co

²Grupo de Estudios para la Remediación y Mitigación de Impactos Negativos al Ambiente - GERMINA
Universidad Nacional de Colombia. Avenida Carrera 50 45-03, Bogotá D.C - Colombia
Tel: (1) 316 5000, correo: dahumada@unal.edu.co

RESUMEN

El aplicativo validaR es una herramienta que permite realizar de manera sencilla numerosos tratamientos estadísticos a datos que típicamente se generan en los procesos de medición química. Con este aplicativo se pretende acompañar a las personas de los laboratorios en su búsqueda por la obtención de conclusiones objetivas a partir de resultados numéricos, de una manera consistente con técnicas estadísticas internacionalmente aceptadas. En este artículo se describen las principales funciones implementadas en validaR, su utilidad en el contexto de procesos como la verificación y validación de métodos químicos cuantitativos y se mencionan algunas de sus perspectivas futuras.

PALABRAS CLAVES

Tratamiento estadístico de datos, aplicativo validaR, aplicación Shiny, identificación de datos anómalos, parámetros de validación.

1. INTRODUCCIÓN

En el análisis químico cuantitativo es imposible controlar por completo todos los factores que influyen sobre el valor de un resultado. Los parámetros que no se pueden controlar o que son desconocidos causan una variación intrínseca en los sistemas a la que se le denomina error aleatorio. El error aleatorio es entonces impredecible y es omnipresente siempre que se realiza una medición de cualquier tipo. Si bien nuestros datos pueden contener la información que es de nuestro particular interés, esta información está siempre distorsionada por la presencia del error aleatorio. Esto es muy relevante porque los resultados que genera un laboratorio con frecuencia influyen decisiones que pueden impactar la salud, el ambiente, la agricultura y muchos otros aspectos que tienen que ver con la actividad humana. En este orden de ideas, es indispensable considerar siempre la existencia del error aleatorio y utilizar herramientas estadísticas que permiten generar conclusiones válidas para tomar decisiones informadas [1].

Muchos laboratorios realizan sus tratamientos estadísticos con programas de hoja de cálculo que pueden ser lo suficientemente buenos para algunos de los fines propuestos, pero que realmente no están diseñados para tal fin y por lo tanto presentan numerosas limitaciones que entorpecen el proceso de organizar, analizar y concluir adecuadamente con los datos [2]. Una alternativa más apropiada es utilizar algún programa estadístico que provea funciones más completas y apropiadas para el análisis de datos, pero en muchas ocasiones esto puede requerir algunos conocimientos básicos de programación y esto puede

representar una limitante para algunos de los laboratorios. Como una alternativa, se presenta el aplicativo **validaR** (pronunciado */validar/*) que es una plataforma que integra varias funcionalidades del muy poderoso lenguaje de programación **R**, diseñado para computación estadística y representaciones gráficas de alta calidad [3].

Desde 2019, en el Grupo de Investigación en Metrología Química y Bioanálisis de la Subdirección de Metrología Química y Biología del Instituto Nacional de Metrología de Colombia (INM) y en el Grupo de Estudios para la Remediación y Mitigación de Impactos Negativos al Ambiente - GERMINA de la Facultad de Ciencias de la Universidad Nacional de Colombia, se ha venido desarrollando este aplicativo con el fin de facilitar el acceso a herramientas estadísticas de alto nivel disponibles en R, pero que de otra forma requerirían conocimientos de programación para su uso. ValidaR funciona como un tablero dinámico que se puede utilizar desde la mayoría de los navegadores de internet accediendo a la dirección web <http://186.155.29.169:3838/validaR/>.

La página de inicio del aplicativo validaR se muestra en la Figura 1. El objetivo de este artículo es dar a conocer las generalidades del aplicativo y poner en contexto la importancia de sus funciones. No se ahonda en el trasfondo teórico de los métodos implementados ni se pretende dar instrucciones detalladas sobre el uso del aplicativo. Para estos fines se recomienda revisar algún libro apropiado de métodos estadísticos [1, 4] y los tutoriales en video que se encuentran disponibles en el canal de YouTube del aplicativo [5], respectivamente.



Figura 1. Página de inicio del aplicativo validaR.

2. GENERALIDADES DEL APLICATIVO

ValidaR es una aplicación interactiva tipo Shiny. Shiny es una librería de R que permite combinar el poder computacional del lenguaje de programación con la interactividad y accesibilidad que ofrecen las aplicaciones web [6]. El resultado es un tablero interactivo que permite manipular y analizar datos sin la necesidad de tener R instalado en el computador ni de saber editar líneas de código fuente. Dos antecedentes importantes de aplicaciones Shiny en el contexto de la metrología son *NIST Uncertainty Machine*, para la estimación de incertidumbres de medición [7] y *NIST Consensus Builder*, para combinar resultados de medición de distintos laboratorios o por distintos métodos de medición [8]. Ambas fueron desarrolladas por miembros del Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST).

El tablero del aplicativo validaR se divide en tres partes: encabezado, panel lateral y panel principal. El panel lateral (en color azul, parte izquierda de la Figura 1) permite navegar entre los distintos módulos del aplicativo. El panel principal ocupa la mayoría de la pantalla y es donde se interactúa en mayor proporción con el aplicativo por medio del ingreso de datos, la configuración de opciones para los módulos y la salida de resultados gráficos y numéricos. Todas las secciones del aplicativo inician con una caja desplegable de instrucciones o de introducción que orientan en la forma de usar las funciones.

El esquema de trabajo general en el aplicativo validaR para el análisis estadístico de datos consta de los siguientes pasos:

1. Generar la matriz para el ingreso de datos e introducir los datos en la tabla.
2. Navegar al módulo de la herramienta estadística de interés, seleccionando la opción que corresponda en el panel lateral.
3. Para usar los módulos, en la región correspondiente del panel principal se debe seleccionar la serie o las series de datos que se vayan a estudiar y se pueden configurar algunas opciones que ajustan el tipo de análisis estadístico que se aplicará.
4. Cuando se presiona el botón de calcular o correr análisis los resultados numéricos aparecen en la

misma página acompañados por lo general de una pequeña descripción de su significado y, ocasionalmente, algunos gráficos que los complementan.

2.1. Ingreso de datos a validaR

Los datos pueden ingresarse manualmente de manera individual o copiando y pegando desde cualquier hoja de cálculo. La matriz de datos almacena todas las cifras significativas ingresadas, pero por defecto solo muestra dos cifras decimales. Al seleccionar una celda con doble clic se puede ver el valor completo del dato que contiene. Si las celdas no llegasen a ser suficientes se debe actualizar el aplicativo y configurar un número de filas y de columnas adecuado, antes de generar la matriz de ingreso de datos.

Los datos se deben organizar en columnas donde cada columna contiene una serie de datos. No se deben incluir etiquetas o nombres para las columnas. Se debe iniciar en la primera fila de cada columna y no se deben dejar columnas vacías a la izquierda. A modo de ejemplo, en la Tabla 1 se muestran datos del porcentaje de recuperación de un analito en un mismo material utilizando procedimientos de extracción con distintos tratamientos. La Figura 2 muestra cómo se verían los mismos datos, ingresados en el aplicativo validaR. Para eliminar un dato de la serie de datos basta con borrar el contenido de la celda correspondiente en la matriz de ingreso de datos.

Trat. 1	Trat. 2	Trat. 3	Trat. 4	Trat. 5
97.374	96.359	100.512	95.955	95.188
98.184	98.975	99.390	95.984	95.838
97.164	99.477	98.379	96.944	95.564
99.595	99.152	96.785	96.821	94.149
98.330	97.389	100.125	86.003	93.020

Tabla 1. Datos experimentales de recuperación (porcentaje) de un analito en un mismo material, utilizando distintos tratamientos para la extracción.

	V1	V2	V3	V4	V5	V6
1	97.37	96.36	100.51	95.96	95.19	
2	98.18	98.98	99.39	95.98	95.84	
3	97.16	99.48	98.38	96.94	95.56	
4	99.60	99.15	96.79	96.82	94.15	
5	98.33	97.39	100.13	86.00	93.02	
6						

Figura 2. Datos de la Tabla 1 cargados en el aplicativo validaR.

3. MÓDULO DE ESTADÍSTICA DESCRIPTIVA

La estadística descriptiva permite resumir la información más relevante de una serie de datos [4]. En el módulo de estadística descriptiva de validaR se

generan distintos resultados que se listan a continuación:

1. Medidas de localización, dispersión y forma.
2. Medidas de posición (cuartiles, percentiles).
3. Gráficos de histograma, puntos apilados y normalidad (cuartil-cuartil).
4. Pruebas de normalidad: Shapiro-Wilk y Kolmogorov-Smirnov (la segunda solo para conjuntos grandes de datos).
5. Pruebas de datos anómalos: Criterio de Dixon y criterios de Grubbs.

Para operar el módulo se debe seleccionar la columna que contiene la variable de interés, escoger un nivel de confianza apropiado para las pruebas del módulo y presionar el botón de calcular.

Los criterios de evaluación de datos anómalos de Grubbs permiten evaluar más de un dato anómalo a la vez en la misma muestra estadística y son los recomendados, por ejemplo, en el análisis de datos de precisión y veracidad de los métodos de medición [9]. En el uso de estas herramientas se debe considerar que cuando uno o más datos son señalados como atípicos se recomienda investigar, en la medida de lo posible, las causas de su desviación anormal, evitando descartarlos únicamente con base en las evidencias estadísticas [10].

4. MÓDULO DE ESTADÍSTICA INFERENCIAL

Con frecuencia aparecen escenarios en los que se debe comparar un descriptor de una muestra estadística (como su promedio o su desviación estándar relativa) contra un valor de referencia, contra el descriptor de otra muestra estadística o dentro de un conjunto de descriptores provenientes de un conjunto de varias muestras estadísticas. En el ejemplo de los datos de recuperación de la Tabla 1 podría resultar de interés comparar la media de uno de los grupos contra un valor ideal de recuperación de 100% (comparar su media muestral contra un valor de referencia), o comparar si dos de los tratamientos dan lugar a extracciones con la misma repetibilidad (comparar sus varianzas) o comparar si hay diferencias en los promedios en al menos dos de grupos en ese conjunto de datos (comparar más de dos medias muestrales simultáneamente). Las posibles diferencias a las que haya lugar deben evaluarse considerando que las entidades que se comparan son afectadas por aleatoriedad. La aleatoriedad puede hacer parecer que dos muestras estadísticas que provienen de una misma población tengan descriptores muestrales diferentes. Una de las estrategias para afrontar este problema es por medio de pruebas de inferencia estadística o de comparación [4, 11].

El módulo de estadística inferencial de validaR se compone de varios sub-módulos de pruebas de comparación.

4.1. Pruebas de comparación de medias

El submódulo de pruebas de comparación de medias permite comparar una media muestral contra un valor de referencia o contra otra media muestral. Se implementaron pruebas t de Student que permiten contrastar hipótesis alternativas unilaterales o bilaterales. En el caso de la comparación de dos muestras estadísticas, es posible considerar las muestras estadísticas como independientes o emparejadas, y aplicar la variante de la prueba t de Student que corresponda. ValidaR detecta automáticamente si las varianzas de los grupos pueden asumirse como iguales o no. También se incluyeron versiones no paramétricas de estas pruebas estadísticas: La prueba de sumas de rangos de Wilcoxon y la prueba de Wilcoxon-Mann-Whitney [11].

La comparación simultánea de varias medias muestrales se realiza en el submódulo de análisis de varianza (ANOVA). Al correr el análisis se ponen a prueba de manera automática los supuestos que tiene el ANOVA (normalidad, homocedasticidad y ausencia de datos anómalos significativos) y se genera la tabla de ANOVA en conjunto con gráficos de diagramas y bigotes (mostrado en la Figura 3) y de residuales, que pueden ayudar a entender mejor los resultados. Cuando los resultados del análisis de varianza resultan estadísticamente significativos, se pueden correr distintas pruebas de comparaciones múltiples de comparación de medias:

1. Diferencia mínima significativa de Fisher.
2. Diferencias honestas significativas de Tukey.
3. Rangos múltiples de Duncan.

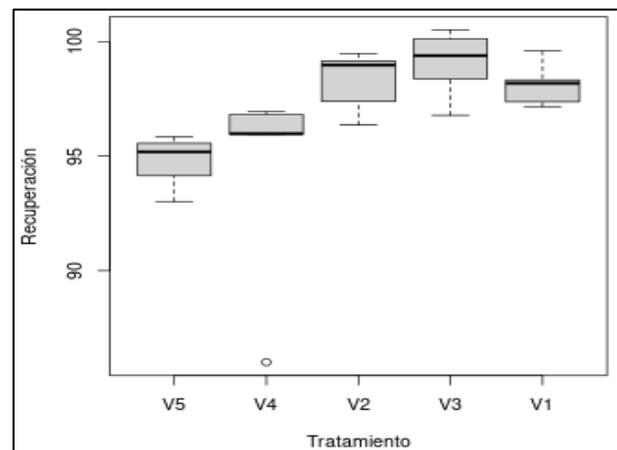


Figura 3. Diagrama de cajas y bigotes generado en el módulo de ANOVA de validaR.

4.2. Pruebas de comparación de varianzas

El submódulo de pruebas de comparación de varianzas contiene distintas herramientas:

1. Prueba χ^2 (pronunciado /ji cuadrado/) para evaluar la dispersión de una muestra estadística contra un

- valor de referencia que puede ser varianza, desviación estándar o coeficiente de variación.
2. Prueba F para evaluar la dispersión de una muestra estadística respecto a la de otra muestra estadística.
 3. Prueba de Cochran para evaluar si una muestra estadística tiene una dispersión atípicamente grande entre un conjunto de varias muestras estadísticas [9].
 4. Pruebas de Levene, de Bartlett y de Hartley para evaluar si las dispersiones de un conjunto de varias muestras estadísticas son diferentes entre sí (heterocedásticas).

Las pruebas χ^2 y F, igual que las pruebas de comparación de medias, permiten configurar hipótesis alternativas unilaterales o bilaterales que pueden afinar el desempeño del análisis estadístico según el tipo de evidencia que se busque con el estudio. Si se desea evaluar diferencias de cualquier tipo (mayor que o menor que) se debe seleccionar una prueba bilateral. Si por el contrario, se desea confirmar que la dispersión de los resultados de medición de un mensurando son mejores que un coeficiente de variación dado (valor de referencia), interesa solo la evidencia que indique que la varianza de la muestra es menor a la correspondiente a ese valor de referencia, y lo conveniente es utilizar una comparación unilateral [12].

4.3. Análisis de covarianza

El submódulo de análisis de covarianza (ANCOVA) permite comparar las medias de dos o más muestras estadísticas en los casos en las que estas son afectadas por otra variable cuyo efecto no es de interés pero que puede afectar negativamente la validez de las conclusiones si no se le considera. El ANCOVA combina las propiedades del ANOVA con las de la regresión lineal y es muy útil en algunos diseños de experimentos empleados en el desarrollo de métodos de medición [13].

5. MÓDULO DE MODELOS DE REGRESIÓN LINEAL

Los modelos de regresión lineal establecen relaciones matemáticas entre al menos dos variables que se pueden denominar explicatoria y respuesta, respectivamente. Estos modelos pueden utilizarse, por ejemplo, para estimar la concentración de especies en muestras analíticas [14], para comparar los resultados de distintas técnicas de medición [15] o para evaluar la estabilidad de materiales de referencia certificados [10].

El método más utilizado para calcular los parámetros de regresión se conoce como mínimos cuadrados ordinarios (OLS) y funciona por medio de la minimización de las distancias entre los valores de la variable respuesta y los valores que predice una línea de tendencia que trata de pasar tan cerca como sea

posible de todos los puntos [4]. El algoritmo OLS tiene unos supuestos que con regularidad son ignorados o que se dan por cumplidos sin una verificación previa [12]. El aplicativo validaR realiza una verificación de la mayoría de estos supuestos.

Algunos estudios en el contexto de las mediciones pueden necesitar otro tipo de algoritmos de estimación de parámetros de regresión lineal que también están implementados en el aplicativo validaR:

1. Mínimos cuadrados ponderados (WLS), para cuando el error de los puntos experimentales no puede asumirse constante en el intervalo de valores considerado. Este tipo de regresión requiere conocimiento de la dispersión o la incertidumbre asociada con cada punto.
2. Mínimos cuadrados por distancia ortogonal (ODR), para cuando el error de los datos en la variable explicatoria no es despreciable respecto al error de los datos en la variable respuesta.
3. Mínimos cuadrados generales (GLS) para cuando se conoce el error de cada punto en ambas variables. Este es el mejor método cuando se comparan los resultados que generan dos técnicas de medición en un intervalo determinado.
4. Regresión no paramétrica de Passing-Bablok, útil cuando el error en las mediciones no se distribuye de manera normal.

El módulo de modelos de regresión lineal tiene un submódulo de interpolación para estimar valores de la variable explicatoria correspondientes a nuevos valores de la variable respuesta. Si la regresión se realizó por medio del algoritmo OLS, el valor interpolado se acompaña del aporte de incertidumbre asociada al uso del modelo de regresión [4].

6. PERSPECTIVAS

El aplicativo validaR se encuentra siempre en evolución y se espera que se incorporen funciones con una aplicabilidad cada vez más directa a los procesos de los laboratorios de análisis químico. Esto incluiría el apoyo en la planeación y análisis de los resultados de los experimentos para evaluar los parámetros de validación de los métodos de análisis químico cuantitativo, como la robustez [16] o los límites de detección [17]. Para esta mejora constante es fundamental la retroalimentación que se reciba por parte de las personas en los laboratorios. Para enviar sugerencias y reportar errores, comedidamente se solicita diligenciar el formulario al que se puede acceder en la penúltima opción del panel lateral.

AGRADECIMIENTOS

La Oficina De Informática Y Desarrollo Tecnológico del INM facilitó los recursos para hospedar el aplicativo en los servidores del Instituto. Los autores agradecen al

Profesor Jesús Ágreda por sus enseñanzas en el lenguaje de programación R y por su apoyo desde el comienzo en la creación de la herramienta validaR.

Descargo de responsabilidad

La herramienta validaR se ofrece tal cual se publica en su página web. Los autores se comprometen a poner esfuerzos para complementarla, corregir posibles errores y actualizarla constantemente, pero no se da garantía de ningún tipo sobre la exactitud de los resultados que se generan ni de la aptitud de estos para un propósito particular. No se garantiza que el funcionamiento del aplicativo será ininterrumpido o libre de errores.

REFERENCIAS

1. Ellison, S. L. R., Farrant, T. J., & Barwick, V. (2009). Practical statistics for the analytical scientist: A bench guide (2nd ed). RSC Publishing.
2. Tanavalee, C., Luksanaprukha, P., & Singhatanadgige, W. (2016). Limitations of using microsoft excel version 2016 (MS Excel 2016) for statistical analysis for medical research. *Clinical Spine Surgery: A Spine Publication*, 29(5), 203-204. <https://doi.org/10.1097/BSD.0000000000000382>
3. R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria. Recuperado el 29 de septiembre de 2021 de <https://www.R-project.org/>
4. Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2nd ed). Wiley-Interscience.
5. Aplicativo validaR [Canal de Youtube]. Recuperado el 29 de septiembre de 2021 de <https://www.youtube.com/channel/UCwklGbFJQVQzVtjhFpo1PSA>.
6. Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Diepert, A., & Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.6.0. Recuperado el 29 de septiembre de 2021 de <https://CRAN.R-project.org/package=shiny>.
7. Lafarge, T. & Possolo, A. (2020). NIST Uncertainty Machine. Recuperado el 29 de septiembre de 2021 de <https://uncertainty.nist.gov/>
8. Koepke, A., Lafarge, T., Possolo, A., & Toman, B. (2020). NIST Consensus Builder. Recuperado el 29 de septiembre de 2021 de <https://consensus.nist.gov/app/nicob>.
9. International Organization for Standardization ISO. (2019). Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. ISO 5725-2:2019. Recuperado el 30 de septiembre de 2021 de <https://www.iso.org/obp/ui/es/#iso:std:iso:5725:-2:ed-2:v1:en>.
10. International Organization for Standardization ISO. (2016). General requirements for the competence of reference material producers. ISO 17034:2016. Recuperado el 30 de septiembre de 2021 de <https://www.iso.org/obp/ui/es/#iso:std:iso:17034:ed-1:v1:en>.
11. Mann, H. B., Whitney, D. R. (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". *Annals of Mathematical Statistics*. 18(1), 50–60. <https://doi.org/doi:10.1214/aoms/1177730491>.
12. Possolo, A., National Research Council Canada, & Sistema Interamericano de Metrología. (2020). Measurement uncertainty: A reintroduction. http://publications.gc.ca/collections/collection_2021/cnrc-nrc/NR16-330-2020-eng.pdf
13. International Organization for Standardization ISO. (2013). Statistics — Vocabulary and symbols — Part 3: Design of experiments Recuperado el 30 de septiembre de 2021 de <https://www.iso.org/obp/ui/#iso:std:iso:3534:-3:ed-3:v1:en>
14. Ahumada, D. A., Soto, L. L., Morales, L. V., & Abella, J. P. (2019). Desarrollo de un material de referencia certificado para análisis elemental de agua potable. *Revista Colombiana de Química*, 48(3), 36-44. <https://doi.org/10.15446/rev.colomb.quim.v48n3.78660>
15. Stöckl, D., Dewitte, K., & Thienpont, L. M. (1998). Validity of linear regression in method comparison studies: Is it limited by the statistical model or the quality of the analytical input data? *Clinical Chemistry*, 44(11), 2340-2346. <https://doi.org/10.1093/clinchem/44.11.2340>
16. Instituto Colombiano de Normas Técnicas y Certificación ICONTEC. (2010). Práctica normalizada para realizar ensayos de robustez para la evaluación estadística de métodos de ensayo. NTC 5826:2010. Recuperado el 30 de septiembre de 2021 de <https://tienda.icontec.org/gp-practica-normalizada-para-realizar-ensayos-de-robustez-ruggedness-para-la-evaluacion-estadistica-de-metodos-de-ensayo-ntc5826-2010.html>
17. Thompson, M., Ellison, S. L. R., & Wood, R. (2002). Harmonized guidelines for single-laboratory validation of methods of analysis (Iupac technical report). *Pure and Applied Chemistry*, 74(5), 835-855. <https://doi.org/10.1351/pac200274050835>