

## REPRODUCIBILIDAD DE LA INVESTIGACIÓN Y EDUCACIÓN ESTADÍSTICA CON R MARKDOWN

HUMBERTO CUEVAS  
Tecnológico Nacional de México

CRISTINA SOLÍS  
Universidad Técnica Nacional, Costa Rica

**RESUMEN:** El objetivo del presente estudio fue utilizar R Markdown como estrategia didáctica para introducir la reproducibilidad en un curso de estadística. Se propuso un caso de estudio que requirió articular habilidades y saberes de cuatro disciplinas. Se encontró que R Markdown es una opción factible de implantar en las aulas universitarias por su potencial para reproducir resultados de investigación en la enseñanza de la estadística, aunque una proporción representativa manifestó renuencia en utilizarlo para otras actividades. Se recomienda efectuar estudios de corte experimental que incluyan el uso de una taxonomía de desempeños, utilizar lenguajes de programación en el análisis estadístico, como medio para desarrollar el pensamiento crítico en detrimento del desiderativo y revisar tecnologías computacionales alternas.

**PALABRAS CLAVE:** *educación estadística, investigación reproducible, R Markdown, tecnología computacional, ética en investigación*

### REPRODUCIBILITY OF RESEARCH AND STATISTICAL EDUCATION WITH R MARKDOWN

**ABSTRACT:** The purpose of this study was to use R Markdown as a teaching strategy to introduce reproducibility in a statistics course. A case study was proposed; in which it was required the articulation of skills and knowledge from four disciplines. It was found that R Markdown is a feasible option to implement in university classrooms due to its potential to reproduce research results in the teaching of statistics, though a representative proportion expressed reluctance to use it in other activities. It is advisable recommended to carry out experimental studies that include the use of a performance taxonomy, use of programming languages in statistical analysis as a way to develop the critical thinking to the detriment of desiderative thinking and review alternative computational technologies.

Recibido: 26 de junio de 2018 • Aceptado: 10 de noviembre de 2018.

KEYWORDS: *statistical education, reproducible research, R Markdown, computational technology, research ethics.*

## 1. *Introducción*

¿En qué medida son precisos los resultados de una investigación? ¿Están disponibles los datos y métodos estadísticos que se usaron? ¿Hasta qué punto es posible rehacer las representaciones gráficas o tabulares? ¿Es posible recrear de manera independiente los hallazgos originales? ¿Siguen teniendo vigencia los datos, su formato y las herramientas tecnológicas empleadas para su análisis? Estas interrogantes son frecuentes entre la comunidad científica, académica y estudiantil cuando examinan reportes de investigación.

Diversas comunidades epistémicas se han pronunciado por la necesidad de establecer estándares, protocolos de trabajo y plataformas tecnológicas, las cuales permitan verificar la confiabilidad y validez de reportes de investigación en el ámbito científico y académico. La reproducción y replicación de este tipo de reportes constituyen alternativas que se han puesto sobre la palestra para su análisis. Los motivos son múltiples, por un lado, destaca el aumento de publicaciones cuyos resultados carecen de evidencia que los sustente y, por otro, la existencia de dudas respecto de procedimientos de acopio, tratamiento de datos, manipulación inapropiada de imágenes, e interpretación deficiente de resultados.

Ioannidis (2005) destacó la preocupación relacionada con la veracidad de los hallazgos de investigación y planteó conjeturas respecto del aumento en la retractación de reportes con resultados falsos o de procedencia dudosa. Destacó el número de estudios realizados con las mismas interrogantes de indagación, los efectos de intereses económicos en centros de investigación públicos y privados, la selección de los estadísticos de prueba utilizados y la búsqueda pragmática de significación estadística, todos ellos como mecanismo de legitimación.

Posteriormente, Fang, Grant & Casadevall (2013), dirigieron un estudio que tuvo como meta examinar las causas principales por las que 2047 artículos de investigación, registrados en la base de datos *PubMed*, atravesaron por un proceso de retractación de sus autores. Encontraron que el 67.4% de retractaciones se debieron a una conducta inadecuada, 43.4% a fraude, 14.2% por duplicar o reciclar estudios previos y un 9.8% por plagiar resultados de otros reportes. El origen geográfico estuvo en función de la causa, no

obstante, aproximadamente el 75% provino de los Estados Unidos de América, República Federal de Alemania, Japón, y la República Popular de China. El estudio de Haruko Obokata (Obokata et al., 2014), sobre células madre, es un ejemplo de manipulación de imágenes para fundamentar hallazgos imposibles de reproducir, el cual fue retirado de manera vergonzosa.

En un estudio efectuado por Baker (2016), se entrevistó a 1576 investigadores de varias disciplinas; Biología (703), Química (106), Tierra y Medio Ambiente (95), Medicina (203), Física e Ingeniería (236), otras (233). La autora del estudio, además, es editora especializada en reportes sobre células madre para la revista *Nature*. Encontró que más del 70% de los entrevistados mencionó que tuvo dificultades para reproducir experimentos y los resultados correspondientes. El 50% indicó que fue incapaz de reproducir sus experimentos en laboratorios propios. Aproximadamente, el 73% de los entrevistados subrayó que al menos la mitad de los estudios en sus disciplinas eran reproducibles. No obstante, los especialistas en Física y Química mostraron más confianza en la capacidad de reproducción.

Patil, Peng, & Leek (2016) revisaron la problemática existente en definir y distinguir el significado de reproducibilidad y replicabilidad desde la perspectiva de la estadística como disciplina. Destacaron el debate actual respecto de la acepción dada al término en áreas como Psicología, Geonómica, Economía, Medio Ambiente y Medicina; además indicaron que las diferencias van más allá de la semántica por sus implicaciones científicas y políticas. Propusieron la configuración de un modelo que integre los componentes de un proceso científico para distinguir las variaciones, protocolos de trabajo, experimentadores, datos acopiados, análisis realizados y características de reportes de investigación.

Por su parte, Pellizzari, Lohr, Blatecky & Creel (2017), examinaron los resultados mencionados por Baker (2016) y externaron la preocupación existente en el ámbito científico. Realizaron un análisis semántico de los términos reproducibilidad y replicabilidad, destacando las connotaciones diferentes del primero en el lenguaje coloquial. Para los autores, reproducir en investigación es obtener los mismos resultados en el estudio original, aun cuando se utilicen métodos de medición, sujetos o condiciones experimentales, que varían en algún grado. Además, señalaron que la replicación es más rigurosa y restrictiva, porque exige usar los mismos métodos, instrumentos, laboratorios, sujetos u objetos, para obtener un resultado idéntico.

Respecto de la acepción del término *reproducibilidad*, Berez-Kroeker et al. (2018) condujeron un estudio cuyo propósito fue examinar su significado en la

investigación lingüística. El estudio fue auspiciado por la Fundación Nacional de Ciencia (NSF) de los EUA. Participaron 41 especialistas en subáreas como sintaxis, fonética, semántica, sociolingüística, tipología, dialectología, lingüística computacional, entre otras. Mencionaron que la diferencia entre la reproducibilidad y la replicabilidad estriba en que la primera es una evolución de la segunda y recomiendan a profesores, investigadores y estudiantes, establecer estándares para evaluar el trabajo con datos, además de reflexionar sobre su reticencia a compartirlos.

Puede notarse que la falsificación, plagio y duplicación de resultados constituyen problemas de gran magnitud en la investigación científica, como se describió líneas atrás. Es posible que lo anterior sea causa de la “presión” por publicar descubrimientos, en detrimento de resultados negativos o replicación de estudios anteriores. Los gobiernos son renuentes a invertir en la replicación o reproducción de estudios efectuados con antelación para verificar la precisión de los resultados; las organizaciones de capital privado son más cautelosas, porque su patrocinio se considera un dispendio. En el caso de las revistas especializadas, la mayoría da preferencia a las contribuciones que presentan resultados originales y, en el mejor de los casos, el envío de bases de datos y código de programación es optativo. Sea cual fuere el caso, el hecho de obtener los componentes de un proceso científico, permite continuar arando los surcos que marca el estudio investigativo.

La reproducibilidad y replicabilidad son atributos irreductibles de la ciencia *per se*; sin embargo, al parecer, su significado y método para implementarlas están en función del modelo educativo en que fueron formados los autores, comunidad epistémica a la que pertenecen y el concepto de ética e integridad científica que profesan.

Para fines prácticos, se puede definir la reproducción como la acción de repetir una investigación, respetando el objetivo, método, protocolo de trabajo, instrumentos de acopio, medición, análisis y contexto originales, con el propósito de obtener resultados idénticos o similares. De ser posible, los sujetos u objetos de estudio deben ser los mismos, o en su defecto, con atributos equivalentes. La replicación es dirigir, de manera independiente, una investigación, respetando el objetivo, método, protocolo de trabajo, instrumentos de acopio, medición y análisis de datos nuevos que permitan recrear los hallazgos y conclusiones. Las distinciones fundamentales se encuentran en que la reproducción demanda el uso de procedimientos y datos originales; la replicación exige que, además de lo anterior, se generen datos nuevos en un entorno que puede ser distinto al original.

No todos los reportes de investigación son de acceso libre ni se pueden reproducir por razones legales y de estructura metodológica, no obstante, es imperativo poner a disposición pública las bases de datos, código de programación, protocolos experimentales y una descripción detallada del método y materiales utilizados. Como señalan Koepsell & Ruiz de Chávez (2015), la publicación de resultados reales es un valor de quien lo hace, los beneficios van más allá de la disciplina en la que se enmarcan, porque la ciencia es un bien social. Lamentablemente, como indican Stern & Elliot (1997), una proporción representativa de investigadores no tienen claro el significado y papel de la ética en sus funciones; como consecuencia, pares académicos, lectores y aprendices parten de la tesis que subraya un comportamiento honesto en sus actividades.

Respecto de la tecnología computacional, es parte inherente en la reproducción de la investigación científica. Destaca el uso de lenguajes de programación y de marcado como R y Markdown, respectivamente; ambos de libre acceso y con apoyo de organizaciones públicas y privadas (Gandrud, 2016; Marwick, Boettiger & Mullen, 2018; Wickham & Grolemund, 2016; Xie, 2015, 2016).

Hoy día, es probable que R sea el lenguaje más usado en la estadística, probabilidad y análisis de datos. Su influencia dentro y fuera de la academia se acrecienta por su potencia de cálculo, capacidad gráfica, variedad de Entornos de Desarrollo Integrado (IDE's) creados *ex profeso* y los más de 11,000 paquetes especializados disponibles. En el caso de Markdown, su baja curva de aprendizaje ha facilitado el desarrollo de implementaciones en otros lenguajes de programación (v.gr. Python, Haskell, Julia).

También, se han desarrollado paquetes con el objetivo de facilitar la reproducibilidad a través de la combinación de texto y código analítico. R Markdown es uno de los más representativos por su capacidad para generar documentos en formatos como pdf, docx y html. Actualmente, existe un grupo de expertos en estadística computacional que continúa con su desarrollo técnico, difusión de sus beneficios en términos de transparencia e impulso en su implantación en la educación.

Es importante subrayar que, en un proceso riguroso de reproducción, es recomendable utilizar software especializado, métodos de análisis estadístico, y elaborar un informe que acate lineamientos metodológicos formales. El uso de R Markdown es una alternativa viable para estos análisis y en la elaboración de reportes dinámicos y reproducibles; porque no es necesario pagar una licencia para su uso, se constituye como una multiplataforma y se distingue por el continuo desarrollo tecnológico.

Aunado a la interacción social y el aprendizaje de la ciencia, la institución escolar constituye un espacio adecuado para introducir la reproducibilidad y promover la honestidad intelectual a través de la aplicación del método científico. El estudio de casos reales, contextuales y multidisciplinares representan recursos factibles y ventajosos para esta introducción, sobre todo porque permiten articular tópicos de asignaturas distintas en un proyecto escolar.

En este caso, la estadística, es más que un compendio de herramientas de análisis o un apéndice de la matemática. Es la disciplina que une el análisis de datos, la programación computacional, los métodos de investigación y la matemática, en lo que hoy se conoce como *ciencia de datos* (Horton, Baumer & Wickham, 2014; Krause, 2016; Zieffler, Garfield & Fry, 2018).

Antaño se daba énfasis en la importancia de la alfabetización estadística como parte esencial en la formación de un ciudadano moderno, notándose en la cantidad de artículos publicados al respecto en las últimas tres décadas. Actualmente, no se discute en qué medida es necesaria esta alfabetización, ni el rol del profesorado, tampoco la importancia de la tecnología computacional; sino que el debate está en cómo lograr la articulación de métodos de trabajo que, además, de eliminar fronteras entre disciplinas, permitan resolver problemas y atender necesidades a través de proyectos integradores.

Así, es más pertinente intentar dar respuesta a interrogantes como las siguientes: ¿Cómo lograr que estudiantes eviten asumir tópicos de sus asignaturas como entes aislados en la solución de un problema? ¿Cómo unir propósitos de asignaturas distintas en un proyecto de trabajo global? ¿Qué características debe reunir un flujo de trabajo que articule *texto-cálculo-análisis de datos-representación gráfica y tabular* en la elaboración de informes reproducibles? ¿Existen diferencias entre estudiantes del sexo masculino y femenino respecto de su desempeño en el uso de R Markdown?

El propósito de este estudio fue describir los resultados obtenidos del uso de R Markdown como estrategia didáctica para introducir la reproducibilidad de la investigación en un curso de estadística. La estrategia integró tópicos estocásticos, programación computacional y un flujo de trabajo desde una perspectiva holista.

## *2. Método y materiales*

Participaron 120 estudiantes universitarios, 63 del género femenino y 57 del masculino. Al momento del estudio, todos eran estudiantes regulares

y cursaban tópicos de probabilidad, estadística y análisis de datos. Se utilizó un diseño de investigación no experimental, porque no se manipularon variables para medir efectos en otras y únicamente interesó acopiar, organizar e interpretar datos en momentos particulares. Se estructuró un caso de estudio que requirió aplicar técnicas estadísticas y programación computacional para su análisis, a través R Markdown. La interacción de los participantes con el caso de estudio consistió en tres etapas:

*Fase 1.* Presentación de un caso de estudio real y contextual que demandó dar respuesta a interrogantes relacionadas con mediciones de longitudes, espesores y profundidades de tornillos de una pulgada. Se hizo énfasis en articular tópicos de tres materias y un taller –estadística inferencial, metrología y normalización, informática y taller de investigación– en un proyecto integrador que permitiera compaginar mediciones, análisis estadístico de datos y programación computacional. Se realizó el planteamiento del problema, interrogantes de investigación, propósito, justificación de su realización, protocolo de trabajo, parámetros de valoración, articulación con los tópicos estadísticos contenidos en el programa, beneficios esperados, método propuesto, instrumentos para acopio y tratamiento de información y calendario de actividades. También, se integraron grupos de cuatro integrantes para la consecución del caso presentado.

*Fase 2.* Se presentaron las características más importantes del lenguaje de programación R y sus aplicaciones en el campo del análisis exploratorio de datos, la estadística descriptiva e inferencial. De forma paralela, se mostraron los atributos principales del lenguaje de marcado Markdown y su enlace con R, a través de R Markdown.

*Fase 3.* En virtud de la necesidad de generar un reporte con los resultados obtenidos, se indicaron los estándares mínimos que debía reunir. Se hizo énfasis en describir los materiales de trabajo, equipo y software especializado. También, se mostró un informe de investigación como referente empírico para que los participantes tuvieran clara la estructura del reporte y las directrices de la norma internacional para el formato de texto, gráficas y tablas estadísticas (ver Anexo 1).

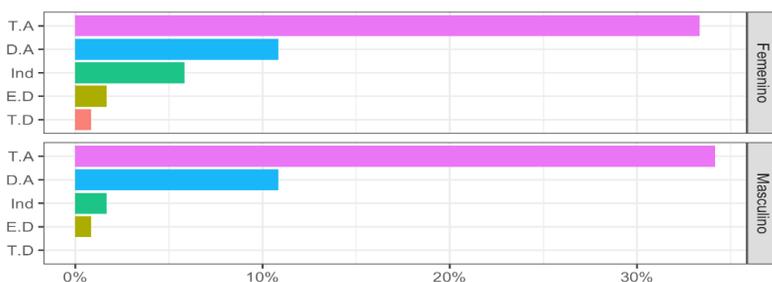
A partir del instrumento desarrollado por Baumer, Cetinkaya-Rundel, Bray, Loi, & Horton (2014), se adaptó un cuestionario de doce ítems; tres para obtener datos generales de los participantes –sexo, edad, materias que cursan– y nueve

para recabar opiniones respecto de usar R Markdown para efectuar procesos de cálculo estadístico, representaciones gráficas, y generación de informes reproducibles (Anexo 2). Las opciones de respuesta fueron las siguientes: Totalmente en desacuerdo (T.D); En desacuerdo (E.D); Ni de acuerdo ni en desacuerdo (Ind); De acuerdo (D.A); Totalmente de acuerdo (T.A). Las estimaciones y contrastes se realizaron con un nivel de confianza del 99% y un nivel de significancia de 0,01; se realizaron pruebas para valorar la diferencia de proporciones por sexo.

### 3. Resultados y discusión

En el ítem 1 se solicitó responder si la sintáxis de R Markdown es fácil de entender. El 67% respondió que estaba *totalmente de acuerdo*, 21,22% *de acuerdo* y sólo el 2,5% *en desacuerdo*. Un participante respondió estar en *total desacuerdo*, sin embargo, sus opiniones fueron contradictorias respecto de otros ítems. Por otra parte, las proporciones de estudiantes que respondieron *totalmente de acuerdo* fueron 0,6349206 y 0,7192982 para el sexo femenino y masculino respectivamente. Con un nivel de confianza del 99%, no se encontraron diferencias significativas [ $-0,3199 \leq |P_1 - P_2| \leq 0,15122$ ]. En la Figura 1 se presenta un panorama global de los resultados.

Figura 1. Resultado global sobre la fácil comprensión de la sintaxis de R Markdown

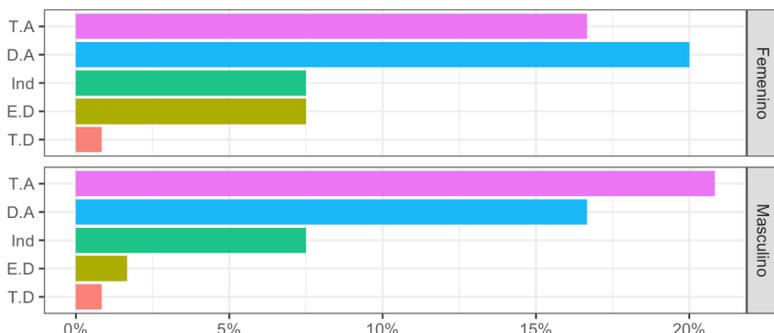


Fuente: Elaboración propia

El ítem 2 se planteó para conocer si cuando tuvieron problemas para compilar un documento de R Markdown, identificaban la causa y realizaban los ajustes necesarios de forma inmediata. El 74,17% señaló que al menos estaba de acuerdo. La Figura 2 permite observar lo anterior. Además, no se encontraron diferencias significativas entre ambos sexos, el *valor*  $-p = 0,3528$

y el intervalo para la diferencia entre proporciones  $[-0,3115 \leq |P_1 - P_2| \leq 0,1294]$ .

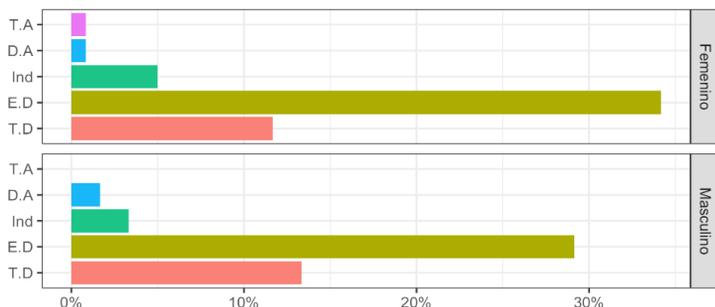
Figura 2. Resolución de problemas al compilar un documento de R Markdown



Fuente: Elaboración propia

Los resultados del ítem 3 indican que el 88,33% de los participantes está en desacuerdo –total o parcial– respecto de preferir utilizar hojas de cálculo y procesadores de palabras para elaborar reportes de investigación o tareas escolares, en detrimento de R Markdown. Menos del 4% indicaron lo contrario, véase la Figura 3. Clasificando los resultados por sexo, los porcentajes fueron 87,30% y 39,47%, no existiendo diferencias significativas porque el *valor*  $-p = 0,9319$  y el intervalo para las diferencias de  $[-0,1888 \leq |P_1 - P_2| \leq 0,1454]$ .

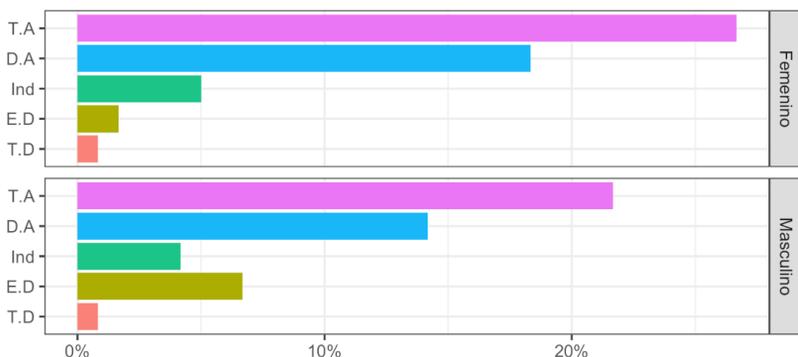
Figura 3. Preferencia en usar R Markdown para reportes de investigación



Fuente: Elaboración propia

Las opiniones vertidas en el ítem 4 confirman lo observado en la práctica cotidiana durante los cursos de estadística. Los participantes señalaron que fue más complejo instalar el lenguaje de programación R y sus paquetes especializados, que aprender su uso. El 100% estuvo De acuerdo con lo anterior, únicamente el 100% discrepó y el 9,17% no estuvo De acuerdo ni en desacuerdo. Los resultados se muestran en la Figura 4. Tampoco, se presentaron diferencias entre proporciones ni se encontró evidencia suficiente para rechazar la hipótesis de igualdad de proporciones con un nivel de significancia del 1%. El intervalo de confianza de 99% [ $-0,0995 \leq |P_1 - P_2| \leq 0,3051$ ] sugiere que  $P_1 - P_2 = 0$ .

Figura 4. Aprendizaje del uso de programación R

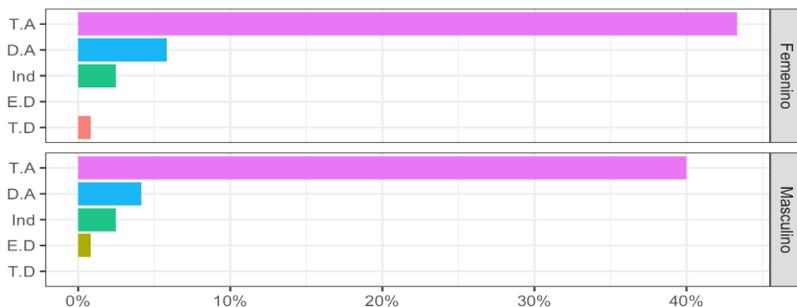


Fuente: Elaboración propia

En relación con las opiniones vertidas para el ítem 5, de los participantes, opinaron que R Markdown hace que sus reportes de trabajo sean más fáciles de leer, entender y reproducir. No se encontraron diferencias entre las opiniones clasificadas por sexo respecto de estar Totalmente de acuerdo con la afirmación planteada – *valor* –  $p = 1$  –. La Figura 5 presenta un panorama general de los resultados.

Sólo una participante opinó estar Totalmente en desacuerdo, sin embargo, dicha opinión contrasta con la emitida en ítems posteriores – 8 y – 9 en el que indica usar R Markdown por su potencia, facilidad de aprendizaje, redactar informes de investigación, además de referir que lo utiliza para presentar reportes en otras asignaturas.

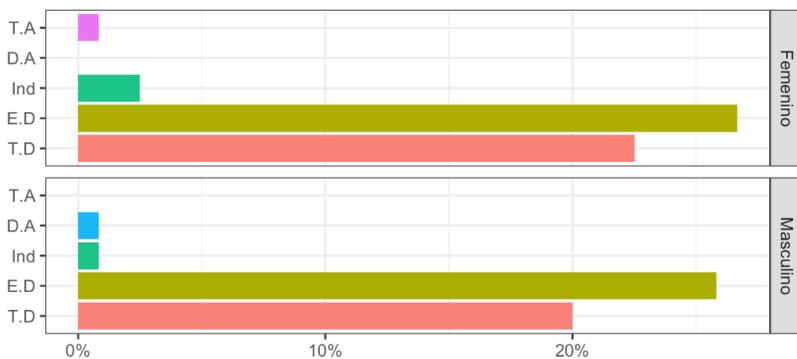
Figura 5. Ventajas al utilizar R Markdown



Fuente: Elaboración propia

El ítem 6 solicitó opinar respecto de su preferencia por continuar efectuando el proceso tradicional de copiar y pegar gráficas, tablas y resultados en procesadores de texto. Destacó el 95%, 114 de los 120 está al menos en desacuerdo con proseguir con esa práctica. La Figura 6 hace constar lo anterior. Al igual que en ítems anteriores, no se presentaron diferencias entre la proporción por sexo – *valor-p* = 0,7691 –, es decir, no se encontró evidencia que permitiera refutar la hipótesis de igualdad de proporciones .

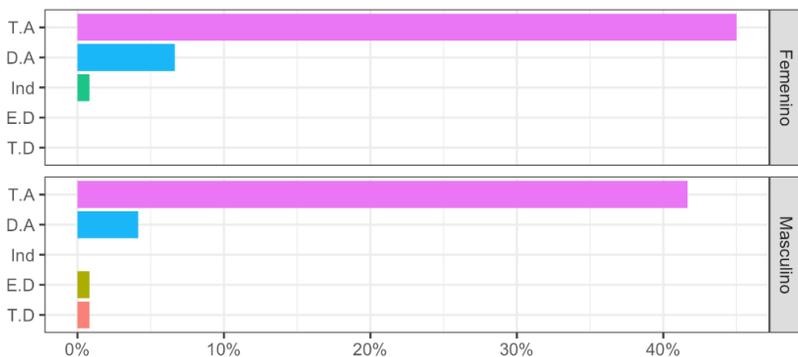
Figura 6. Preferencia al utilizar un proceso tradicional en procesadores de texto



Fuente: Elaboración propia

En el ítem 7 se planteó que R Markdown es muy potente para realizar cálculo estadístico, crear gráficas y tablas, fácil de aprender y redactar informes de investigación. La mayoría (97,5%) de los participantes opinaron estar De acuerdo o Totalmente de acuerdo. Únicamente dos participantes del sexo masculino manifestaron estar En desacuerdo o totalmente en desacuerdo. Fue en este ítem que se presentó la opinión más favorable a R Markdown respecto de su versatilidad entre los nueve ítems que integraron el cuestionario. La Figura 7 ilustra lo anterior.

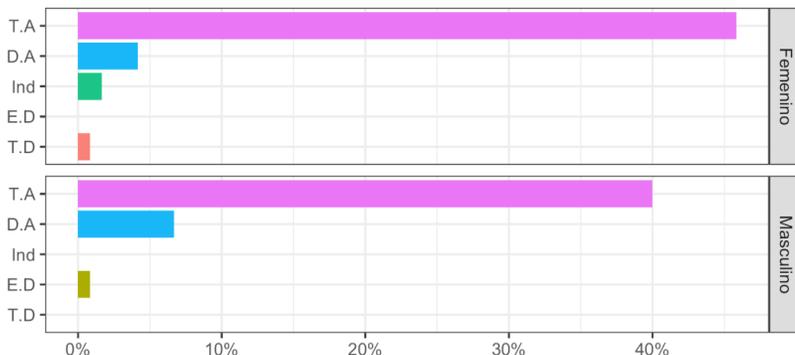
Figura 7. Reconocimiento de la potencia de R Markdown para realizar cálculos estadísticos



Fuente: Elaboración propia

Asimismo, se solicitó opinión respecto de si al inicio del curso fue difícil entender la lógica de trabajo para mezclar código en la elaboración de informes. Alrededor del 86% estuvo Totalmente de acuerdo y un 11% De acuerdo. Las cifras también hacen alusión que después de trabajar durante 15 semanas, los participantes superaron la curva de aprendizaje en la aplicación de R Markdown al proceso de solución del proyecto de trabajo. Los resultados se pueden observar en la Figura 8.

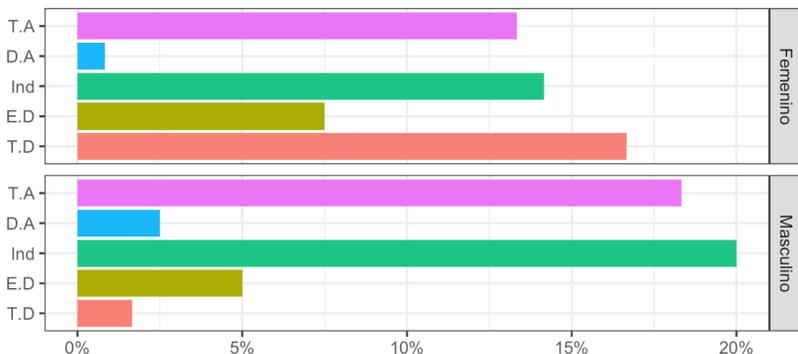
Figura 8. Dificultad de mezclar código en la elaboración de informes



Fuente: Elaboración propia

El último ítem pidió información respecto del uso de R Markdown para presentar reportes en otras materias. Se halló que únicamente el 31,67% estuvo Totalmente de acuerdo, 3,33% De acuerdo, un 34,17% manifestó indecisión, es decir, respondió Ni de acuerdo ni en desacuerdo. El 12,50% mencionó estar En desacuerdo y el 18,33% opinó estar En total desacuerdo. Respecto del último resultado, se encontraron diferencias altamente significativas entre las proporciones de participantes por sexo – *valor - p* = 0,0001728 –. Así, con un nivel de significancia del no se encontró evidencia suficiente que permita validar la hipótesis referente a la igualdad de proporciones; lo más probable es que difieran entre  $[0,31746 \leq | P_1 - P_2 \leq 0,3508772]$ . Los resultados también señalan que prevaleció la indefinición y el desacuerdo en usar R Markdown para presentar reportes en otras materias, especialmente entre los varones. En la Figura 9 se puede notar la diferencia de opiniones clasificada por sexo.

Figura 9. Utilización de R Markdown para elaborar reportes en otras materias



Fuente: Elaboración propia

Es importante apuntar que, de los 30 grupos formados, 27 culminaron el estudio de caso, los restantes abandonaron la actividad antes de concluir la cuarta semana de trabajo. De los grupos que terminaron y presentaron su reporte, 19 dieron respuesta acertada a todas las interrogantes planteadas. Entre las fallas o limitantes encontradas en los reportes de los grupos restantes destacan las siguientes:

1. Uso de representaciones gráficas y tabulares poco apropiadas al tipo de datos.
2. Selección errónea de estadísticos de prueba.
3. Deficiencia en la estructuración de bases de datos.
4. Errores de compilación.

Las primeras dos están dentro del dominio de la estadística y análisis de datos, las restantes en el de programación computacional. En el caso de las representaciones gráficas, dos grupos utilizaron figuras circulares para mostrar datos cuantitativos continuos, con las consecuencias que implica en su lectura e interpretación posterior. En el caso de los estadísticos de prueba, un grupo tuvo dificultad para escoger entre la prueba  $z$  y  $t$  - *student*. Se les entrevistó por separado para indagar las causas; a continuación, se describe un extracto del diálogo sostenido:

- Profesor: ¿Cuáles son los requisitos para usar el estadístico de prueba  $z$ ?
- Estudiantes:
  - Siempre que se conozca la  $\sigma$  o  $\sigma^2$
  - Qué los datos se distribuyan normalmente. Una distribución gráfica de valores con una  $\mu \mp 3\sigma$
  - Muestras grandes  $n \geq 30$ .
- Profesor: En relación con la prueba  $t$  – *student*, ¿cuándo se recomienda su uso?
- Estudiantes:
  - Cuando el tamaño de muestra es pequeño  $n < 30$ .
  - Siempre que se desconozca el valor de la  $\sigma$  o  $\sigma^2$ .
- Profesor: Les plantearé tres interrogantes. Reflexionen y argumenten sus respuestas:
  - ¿Qué estadístico utilizarían si su tamaño de muestra es  $n = 20$  y conocen la  $\sigma$ ?
  - ¿Qué estadístico sugieren usar si el tamaño de muestra es  $n = 50$  y desconocen la  $\sigma^2$ ?
  - ¿Y si  $n = 30$ ?
- Estudiantes:
  - Si  $n = 20$  y conocemos la  $\sigma$  utilizamos la prueba  $t$  – *student*. El argumento es que el tamaño de muestra es menor que 30.
  - No sabemos la respuesta para la segunda pregunta. Estamos confundidos: ¿Debemos calcular la desviación estándar de la muestra y luego suponer que es para utilizar la prueba  $z$ ? o ¿cómo se podría usar  $t$  – *student* si el tamaño de muestra es mayor que 30? Se supone que debemos utilizar  $z$ .
  - Los resultados deben ser aproximadamente iguales. Se debe utilizar  $z$  de preferencia en lugar de  $t$  – *student*.

En el extracto, se observaron las dudas de la elección del estadístico de prueba. Al parecer les queda claro que la prueba  $z$  es recomendable cuando la  $\sigma^2$  es conocida y las muestras son grandes  $n \geq 30$ . Es probable que las dudas que expresaron en la segunda pregunta se debieran a que traían copia de una tabla  $t$  – *student* con valores para más de 100 grados de libertad (gl), es decir, dudaron al observar que dicha tabla excediera los 29 gl típicos.

Es importante subrayar que durante las sesiones de trabajo y el análisis del informe de investigación –referente empírico– presentado al inicio del curso (ver Anexo 1), se examinaron las características más representativas de estas

pruebas y sus diferencias. Así, se probó por ejemplo que cuando los tamaños de muestras son muy grandes, se incrementa la probabilidad de cometer un error tipo II (aceptar una  $H_0$  cuando debiera rechazarse); sin embargo, también se revisó un caso en que se conocía la  $\sigma^2$ , pero existían dudas respecto de su vigencia y se recomendaba calcular  $s$ . Aún cuando el fenómeno se presentó únicamente en un grupo, quizá deba modificarse la estrategia didáctica utilizada durante la selección de los estadísticos de prueba.

Por otra parte, las fallas o limitantes en la programación computacional se presentaron por errores de sintaxis, configuración inadecuada de la codificación (se utilizó UTF-8) y formato usado en bases de datos (se usó separación por comas *csv*).

#### 4. Conclusiones y recomendaciones

La práctica científica atraviesa una crisis de credibilidad. La tasa de artículos que se retiran después de su publicación es alarmante. La retractación obligada no discrimina disciplina alguna ni perfil de comunidades epistémicas. Es una crisis global que se acrecienta y refleja en la falsificación y duplicación de resultados. Es probable que lo anterior sea producto de la “presión” por publicar en medios de alta especialización e impacto internacional que, en el mejor de los casos, ayude al investigador a mejorar sus ingresos económicos, recibir distinciones y tener mayor seguridad laboral. Como se indicó líneas atrás, dichos medios no son proclives a publicar resultados que reproduzcan estudios anteriores ni que tengan una connotación negativa. A lo anterior se suma la renuencia de los gobiernos y empresas privadas a financiar la replicación y reproducción de investigaciones con cierto grado de controversia.

Se acepta que la reproducibilidad y replicabilidad de la investigación son parte inherente de la ciencia. No obstante, continúan existiendo divergencias en la definición de los términos. También, está claro que la acepción de la ética en la investigación es diferente entre comunidades epistémicas, pero eso no obsta para que se establezcan estándares irreductibles de cumplimiento que permitan promover la honestidad y transparencia intelectual; sin estándares, la sociedad seguirá condenada a asumir la tesis que sostiene un comportamiento honesto de comunidades epistémicas y sus integrantes. La promoción de la reproducción y replicación de la investigación es insuficiente para abatir la crisis de credibilidad. Es necesario el saber qué, cómo, cuándo y dónde hacerlo. La tecnología computacional y la educación estadística constituyen parte

toral en la formación de individuos con habilidades en solucionar problemas y atender necesidades de su entorno.

En este artículo se presentaron los resultados de un estudio exploratorio que tuvo como objetivo utilizar R Markdown como estrategia didáctica para introducir la reproducibilidad en el aula. La estrategia consistió en estudiar un caso que demandó el uso y articulación de habilidades y saberes que desarrollaban y alcanzaban de manera paralela en tres materias y un taller –Estadística Inferencial, Metrología y Normalización, Informática, Taller de Investigación– en un semestre de 16 semanas de duración. La estrategia didáctica tuvo etapas de razonamiento inductivo y deductivo en el mismo proceso. Deductivo porque se partió de un problema real, además de que la observación y medición de objetos tangibles a la luz de estándares teóricos permitió extraer conclusiones. Inductivo porque las suposiciones iniciales contenidas en el problema se fueron enriqueciendo con evidencia objetiva conforme transcurría el análisis del caso, se documentaban las observaciones y se generaban hipótesis susceptibles de ser probadas empíricamente. Así, el flujo de trabajo fue un ir y venir explorando datos, creando, eliminando y ajustando hipótesis.

Los participantes indicaron que la curva de aprendizaje para R Markdown es corta. En su mayoría estuvieron de acuerdo en que es muy potente para efectuar cálculos estadísticos, elaboración de gráficas y tablas. Destacaron las ventajas de tener un flujo de trabajo que permita generar informes de investigación a través de un documento único que incluya todo el texto y código de programación necesarios; el proceso tradicional de copiar y pegar entre aplicaciones recibió poco apoyo. También, experimentaron la pertinencia de reproducir estudios de corte cuantitativo como mecanismo para transparentar los resultados y subrayaron que aun cuando tuvieron dificultades al inicio para entender la lógica de trabajo en la mezcla de código, paulatinamente lo entendieron.

Es importante destacar que, salvo el último ítem, no se encontró evidencia suficiente para refutar las hipótesis de igualdad de proporciones entre estudiantes del sexo femenino y masculino. Las estudiantes se manifestaron más renuentes en adoptar la programación computacional con R Markdown; el dato no sorprende porque en sondeos informales, ellas prefieren usar paquetes informáticos, y hojas de cálculo y procesadores de palabras para sus reportes. Debe destacarse que únicamente el 35% utilizó R Markdown para elaborar informes en otras materias. Un porcentaje similar manifestó

indecisión y, aproximadamente, el 31% prácticamente no lo usó. Las causas pueden responder a hipótesis diversas, a saber, [a] evitar las dificultades inherentes al proceso de instalación de los lenguajes y los paquetes especializados; [b] lógica de trabajo diferente a la que están acostumbrados; [c] poca o nula promoción de las virtudes de la programación computacional en el análisis e interpretación de datos por parte de la comunidad de profesores; [d] estrategia didáctica deficiente por parte del conductor del estudio.

De los grupos de trabajo que se formaron, 27 culminaron el estudio de caso y tres lo abandonaron. De los 27, 19 cumplieron con los estándares mínimos de cumplimiento sin falla alguna. Entre los ocho grupos restantes, se encontraron fallas relacionadas con el dominio disciplinar de la estadística y programación computacional. Después de realizar una entrevista semiestructurada con el equipo que más fallos presentó, se dedujo que es altamente probable que deba ajustarse la didáctica y la actuación del profesor, es decir, por las características de las fallas, no se le puede atribuir el no cumplimiento a los estudiantes. El 90% de los participantes culminaron el estudio. El 63% efectuó análisis y generó un reporte de trabajo coherente y preciso en sus cálculos y representaciones tabulares y gráficas.

La última cifra es un porcentaje aceptable, sin embargo, es recomendable realizar modificaciones en lo siguiente:

*Uso de laboratorio.* Preferir el uso de equipo computacional de los laboratorios de cómputo en lugar de computadoras personales. Es preferible que el profesorado instale y configure el software con antelación. El lenguaje de programación R y el paquete R Markdown son de uso libre, código abierto y multiplataforma, lo que les confiere atributos importantes como instalación sobre sistemas operativos diversos; sin embargo, la diversidad de equipos de cómputo, pueden obstaculizar su instalación, ejecución y actualización.

*Periodo.* Se sugiere invertir al menos ocho horas en capacitar a los estudiantes en el uso del lenguaje de programación R. El propósito fundamental debe ser el análisis del caso, aplicar habilidades y saberes multidisciplinares para dar respuesta a interrogantes concretas y generar, probar, reformular o refutar hipótesis. La solución de problemas técnicos siempre estará presente y coadyuva al proceso formativo del futuro profesional, pero no es lo más importante.

*Estudios experimentales.* Es recomendable planear estudios experimentales que permitan tener un control mayor de los factores involucrados en un

proceso de enseñanza y aprendizaje. No siempre es posible seleccionar asignar sujetos de estudio a grupos de control y experimentales, porque dichos grupos ya están formados. No obstante, utilizar diseños cuasiexperimentales pueden estar al alcance del profesorado.

*Uso de taxonomías y estándares mínimos.* El uso de variables categóricas es una práctica valiosa en la investigación educativa porque permite examinar supuestos y describir realidades in situ, además de coadyuvar en establecer uno o más puntos de partida para otro tipo de estudios. Sin embargo, no siempre permiten dar respuestas a interrogantes que exigen respuestas que van más allá de una opinión. ¿Cómo puede clasificarse el grado de dominio disciplinar –en términos de desempeño– de un estudiante? ¿Qué atributos debe reunir un estudiante que cumplió con lo irreductible en un tópico particular? ¿Qué debe saber y ser capaz de hacer para lograr una valoración de suficiencia? Se recomienda establecer estándares confeccionar y validar de taxonomías de desempeño que coadyuven en conocer las fortalezas y deficiencias de los estudiantes para apoyarlos en alcanzar niveles de desempeño superiores.

*Uso de tecnologías alternas.* R y Markdown tienen una aceptación amplia en muchas comunidades, pero existen otras tecnologías que están a la par en términos tecnológicos. Python como lenguaje multipropósito, se ha convertido en una alternativa factible para introducir la reproducción de la investigación en el ámbito educativo. Al igual que R, es de uso libre y permite efectuar articular cálculos y texto cálculo estadístico en un documento. En Kitzes, Turek & Deniz (2017) se pueden analizar casos de estudio en distintas disciplinas; en Yaniv, Lowekamp, Johnson, & Beare (2018) se muestra un ejemplo del uso de Python en un estudio colaborativo en un estudio computacional de alta complejidad; Boscoe, Pasquetto, Golshan, & Borgman (2018) muestran el potencial de Python-Jupyter en la reproducibilidad de la investigación en el campo de la Astronomía; subrayan la variedad de usos de los cuadernos digitales y recomiendan su utilización.

Este reporte forma parte de un estudio longitudinal que tiene como objetivo introducir la reproducibilidad de la investigación en instituciones universitarias a través de una metodología sustentada en la investigación dirigida y solución de problemas reales y contextuales.

REFERENCIAS

- BAKER, M. (2016). Is there a reproducibility crisis? *Nature*, 533, 452-454. DOI <https://doi.org/10.1038/533452a>
- BAUMER, B., CETINKAYA-RUNDEL, M., BRAY, A., LOI, L., & HORTON, N. (2014). R Markdown: Integrating A Reproducible Analysis Tool Into Introductory Statistics. *Technology Innovations In Statistics Education*, 8(1). Tomado de <https://escholarship.org/uc/item/90b2f5xh>
- BEREZ-KROEKER, A., GAWNE, L., KUNG, S., KELLY, B., HESTON, T., HOLTON, G. & WOODBURY, A. (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, 56(1), 1-18. DOI <https://doi.org/10.1515/ling-2017-0032>
- BOSCOE, B., PASQUETTO, I., GOLSHAN, M. & BORGMAN, C. (2018). Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. *arXiv preprint arXiv:1804.05492*.
- FANG, F., GRANT, R. & CASADEVALL, A. (2013). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 110(3), 1138-1143. DOI <https://doi.org/10.1073/pnas.1220833110>
- GANDRUD, C. (2016). *Reproducible research with R and RStudio*. Chapman; Hall/CRC.
- HORTON, N., BAUMER, B. & WICKHAM, H. (2014). Teaching precursors to data science in introductory and second courses in statistics. *arXiv preprint arXiv:1401.3269*.
- IOANNIDIS, J. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), 696-701. DOI <https://doi.org/10.1371/journal.pmed.0020124>
- KITZES, J., TUREK, D. & DENIZ, F. (2017). *The practice of reproducible research: case studies and lessons from the data-intensive sciences*. Univ of California Press.
- KOEPSSELL, D. & RUIZ DE CHÁVEZ, M. (2015). Ética de la investigación, Integridad Científica. (C. N. de Bioética, Ed.) (p. 180). México: CONACYT. DOI <http://www.cooperacionib.org/libro-etica-de-la-Investigacion-gratuito.pdf>
- KRAUSE, A. (2016). Reproducible Research in Real Estate: A Review and an Example. *Journal of Real Estate Practice and Education*, 19(1), 69-85.
- MARWICK, B., BOETTIGER, C. & MULLEN, L. (2018). Packaging data analytical work reproducibly using R (and friends). *The American Statistician*, 72(1), 80-88.
- OBOKATA, H., SASAI, Y., NIWA, H., KADOTA, M., ANDRABI, M., TAKATA, N. & VACANTI, C. (2014). Bidirectional developmental potential in reprogrammed cells with acquired pluripotency. *Nature*, 505(7485), 676.
- PATIL, P., PENG, R., & LEEK, J. (2016). A statistical definition for reproducibility and replicability. *Psychological Science*, 351, 1037-1037.
- PELLIZZARI, E., LOHR, K., BLATECKY, A. & CREEL, D. (2017). *Reproducibility: A Primer on Semantics and Implications for Research* (p. 78). NC: RTI Press. DOI <https://doi.org/https://doi.org/10.3768/rtipress.2017.bk.0020.1708>

- STERN, J. & ELLIOT, D. (1997). *The Ethics of Scientific Research* (p. 75). EUA: University Press.
- WICKHAM, H. & GROLEMUND, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media, Inc.
- XIE, Y. (2015). *Dynamic Documents with R and knitr* (Vol. 29). CRC Press.
- XIE, Y. (2016). *Bookdown: Authoring Books and Technical Documents with R Markdown*. CRC Press.
- YANIV, Z., LOWEKAMP, B., JOHNSON, H. & BEARE, R. (2018). Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research. *Journal of digital imaging*, 31(3), 290-303.
- ZIEFFLER, A., GARFIELD, J. & FRY, E. (2018). What Is Statistics Education? En D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 37-70). Cham: Springer International Publishing. DOI [https://doi.org/10.1007/978-3-319-66195-7\\_2](https://doi.org/10.1007/978-3-319-66195-7_2)

#### AUTORES DEL ARTÍCULO

HUMBERTO CUEVAS, es catedrático del Tecnológico Nacional de México. Monterrey-México. Profesor e Investigador adscrito al Departamento de Ciencias Básicas del Instituto Tecnológico de Chihuahua II. Doctor en Educación por la Facultad de Filosofía y Letras de la Universidad Autónoma de Chihuahua, Maestro en Ciencias en Enseñanza de la Ciencia y Especialista en Docencia por el Centro Interdisciplinario de Investigación y Docencia en Educación Técnica, CIIDET, e Ingeniero Industrial por el Instituto Tecnológico de Delicias. Miembro del Sistema Nacional de Investigadores y Perfil PROMEP. Sus líneas de trabajo son: educación estadística y análisis de datos; formación del profesorado de estadística; aplicación de tecnología informática en educación. Ha dictado conferencias y publicado trabajos de investigación en foros y medios nacionales e internacionales.

Redalyc: [redalyc.org/autor.oa?id=21789](http://redalyc.org/autor.oa?id=21789)

Orcid: [orcid.org/0000-0003-4142-8161](http://orcid.org/0000-0003-4142-8161)

Correo electrónico: [jesus.humberto.cuevas@outlook.com](mailto:jesus.humberto.cuevas@outlook.com)

CRISTINA SOLÍS, es docente de la Universidad Técnica Nacional, Costa Rica. Licenciada en Ciencias de La educación con énfasis en la enseñanza del Español. Universidad de las Ciencias y de las Artes de Costa Rica.

Correo electrónico: [csolis@utn.ac.cr](mailto:csolis@utn.ac.cr)



## ANEXO 1. INFORME DE INVESTIGACIÓN – REFERENTE EMPÍRICO

### Diámetro de piezas metálicas. Un estudio comparativo

ISABEL MOREIRA\*

Tatsuyoshi Motors, Uruguay  
Isa-Bel@Tatsu.com.uy

#### Resumen

Se presenta un caso de estudio relacionado con medidas del diámetro de piezas de motores automotrices. Se efectuó una revisión de dos lotes con producto terminado de dos compañías. El objetivo fue comparar dos muestras representativas de ambas compañías y examinar el grado de cumplimiento con los estándares requeridos [ $\mu = 20, 25\text{mm}$  y tolerancia de  $\pm 2, 25\text{mm}$ ]. Se realizó un análisis exploratorio de datos y pruebas de estadística inferencial con un nivel de confianza del 99%. No existieron diferencias significativas entre las medidas promedio de ambas muestras, sin embargo, en una de ellas se encontró que al menos una pieza no cumplió con las tolerancias y se rechazó.

Palabras clave: estudio comparativo, análisis exploratorio de datos, mediciones

#### Abstract

A case study is presented related to measurements of the diameter of automotive engine parts. A review of two lots with two companies finished product was made. The aim was to compare two representative samples of both companies and examine the degree of compliance with the required standards [ $\mu = 20, 25\text{mm}$  and tolerance of  $\pm 2.25\text{mm}$ ]. An exploratory data analysis and inferential statistics tests were performed with a 99% of confidence level. There were no significant differences between the average measures of both samples, however, one was found that at least one piece did not meet the tolerances and rejected.

Keywords: comparative study, exploratory data analysis, measurements

## 1 Introducción

El Departamento de Ingeniería de una Compañía Manufacturera de Motores Automotrices, organizó un concurso abierto para seleccionar proveedores de piezas metálicas. En la primera etapa, se inscribieron cinco compañías que presentaron un plan de negocios; se examinó el cumplimiento irrestricto de los lineamientos y se descartaron tres por incumplimiento. Las compañías que cumplieron los requisitos fueron: *Componentes Metálicos*, y *Mecatronics United*.

Se requirió seleccionar un proveedor, por tanto, se les solicitó ingresar a sus plantas de manufactura para examinar muestras de producto terminado y dar respuesta a las interrogantes siguientes: [1] ¿Cuál es el diámetro promedio real de cada lote?, [2] ¿En qué grado cumplen con las especificaciones de medida?, [3] ¿En qué medida existen diferencias entre los diámetros promedio entre ambos lotes?, [4] ¿Cuáles son las diferencias entre las variaciones?, [5] ¿Cuál de los dos proveedores produce menos piezas defectuosas?, y [6] ¿Cuál proveedor debe seleccionarse para establecer un contrato? Previamente se les indicó que la medida del diámetro promedio es  $\mu = 20, 25\text{ cm}$ , con tolerancia de  $\pm 2, 25\text{ mm}$ . El propósito fue efectuar un análisis exploratorio y comparativo de las medidas del diámetro solicitadas para las piezas, a la luz del estándar establecido. La realización del estudio fue pertinente porque las piezas forman parte del núcleo del motor. La existencia de no conformancias en las especificaciones

puede detener las líneas de producción de la compañía, con los costos asociados que eso implica por las horas-hombre perdidas, incumplimiento en las fechas de entrega al cliente final y la publicidad negativa.

## 2 Método

La población bajo estudio consistió en dos lotes de producto terminado, uno por cada compañía participante. En presencia de dos representantes del Departamento de Calidad, se seleccionó una muestra de 20 piezas al azar de cada lote. Se midieron los diámetros y se confeccionó una base de datos con formato separado por comas (csv). Los instrumentos utilizados para las mediciones fueron cuatro calibradores vernier digitales de marca *Triper*, cuyas tolerancias fueron  $0,00''/0,05\text{mm}$ , capacidad máxima de  $6''/150\text{mm}$  y mínima de  $0,005''/0,01\text{mm}$ . Para realizar los análisis estadísticos y la generación del reporte escrito, se utilizó el Lenguaje de Programación R 3.4 y el Entorno de Desarrollo Integrado (IDE) RStudio 1.0. Se usó el paquete *stargazer* creado por Hlavac (2015) para presentar un resumen estadístico.

El procedimiento seguido consistió en medir cada pieza en tres ocasiones, calcular el promedio y su desviación estándar. Todas las mediciones fueron realizadas por la misma persona, con el objetivo de evitar la generación de sesgos. La base de datos se integró con las medidas promedio y se procedió a realizar la prueba de normalidad desarrollada en 1965 por Samuel Shapiro y Mar-

\*Jefa del Departamento de Confiabilidad

tin Wilk (Shapiro y Wilk, 1965); la prueba es estadísticamente muy potente cuando se trabaja con muestras pequeñas  $n < 30$ . Posteriormente se realizó un análisis exploratorio de datos.

No obstante, se presentó mayor dispersión en las medidas de Componentes Metálicos  $s_1 = 0,750$  versus  $s_2 = 1,151$  (Vease Figura 2).

### 3 Resultados y discusión

La prueba de normalidad de Shapiro–Wilk arrojó un *valor*  $-p = 0,3474404$  para las medidas de la compañía Componentes Metálicos y  $0,2220737$  para Mecatronics. No se encontró evidencia empírica para rechazar la hipótesis nula, los valores de ambos lotes se distribuyen normalmente. En la Figura 1 se presenta una gráfica cuantilar en que se comparan los valores observados contra los teóricos.

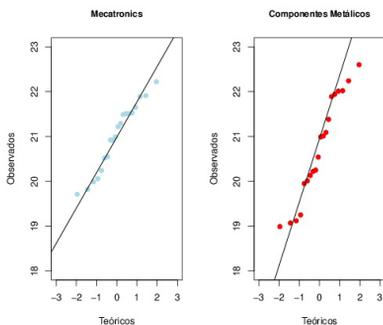


Figura 1: Comparación de cuantiles observados y teóricos

Se realizó un análisis exploratorio de datos (AED) para cada muestra. Se encontró que las medias de los diámetros para las compañías Mecatronics y Componentes Metálicos fueron  $\bar{x}_1 = 20,997mm$  y  $\bar{x}_2 = 20,735$  respectivamente (vease Tabla 1). Los promedios están dentro de las tolerancias establecidas por la empresa y sus desviaciones estándares fueron pequeñas  $s_1$  y  $s_2$  menores a  $3s$  de la media. No se presentaron valores atípicos porque no se excedió el supuesto:  $x > (Q_3 + 3iqr)$  es decir, el valor es mayor que el resultado ubicado en el cuartil tres más el producto de tres veces el rango intercuartílico.

Tabla 1: Resumen estadístico

Statistic	Mecatronics	Componentes Metálicos
N	20	20
Mean	20.997	20.735
St. Dev.	0.750	1.151
Min	19.710	18.990
Pctl(25)	20.450	19.995
Median	21.105	20.765
Pctl(75)	21.515	21.902
Max	22.220	22.600

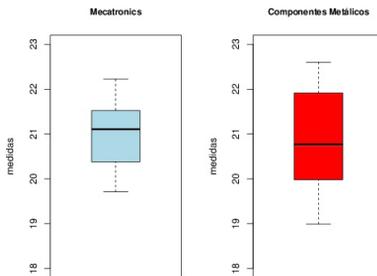


Figura 2: Comparación de cuantiles observados y teóricos

Con un nivel de confianza de 99% se encontró que el promedio real de los diámetros fue un valor entre  $19,99875 < \mu < 21,47125$ . En el caso de Mecatronics, el parámetro  $\mu$  se encontro entre  $20,517119 < \mu < 21,47125$ . En virtud de que se desconocían las varianzas poblacionales, se realizó una prueba al respecto. Se constató con un nivel de confianza de 99%, el cociente de las varianzas de las dos muestras fue  $0,6861063 < \sigma < 8,0802119$  y se asumió que las varianzas fueron idénticas. Así, se calculó un intervalo de confianza para examinar si existían diferencias entre las medidas promedio entre ambas muestras. No se encontraron diferencias  $[-1,0949071 < \mu_1 - \mu_2 < 0,5709071]$ .

### 4 Conclusiones

El diámetro real de cada lote se encuentra dentro de las tolerancias permitidas, ambas empresas cumplen con las especificaciones de medida que se solicitaron [ $\mu = 20,25\text{ cm}$ , con tolerancia de  $\pm 2,25\text{ mm}$ ]. No hubo diferencias significativas entre los diámetros promedio ni en la variación entre ambos lotes. No obstante, la compañía Mecatronics mostró una dispersión menor en las medidas; aparentemente su proceso esta más controlado. Se recomienda firmar un contrato con Mecatronics y establecer pautas relativas a instrumentar un proceso de mejora continua para reducir aún más la variación y producir piezas más homogéneas.

### Referencias

Hlavac, M. (2015). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.  
 Shapiro, S. y Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

## ANEXO 2. INSTRUMENTO PARA ACOPIO DE OPINIONES

### CUESTIONARIO

**Presentación:** El propósito de esta encuesta es reunir información que permita conocer su opinión respecto del uso de tecnología computacional en el tratamiento de tópicos estocásticos. La información será tratada en forma confidencial.

**Instrucciones:** La encuesta consta de 12 ítems. Con excepción de los tres primeros, responda con un valor numérico (1 al 5) a partir de la clasificación siguiente:

Valor	Equivalencia (ocasiones)
-------	--------------------------

1	Totalmente en desacuerdo
2	En desacuerdo: Menos del 40%
3	Ni de acuerdo ni en desacuerdo Entre el 40% y 60%
4	De acuerdo Más del 60% y menos del 100%
5	Totalmente de acuerdo 100%

A. Sexo: [ ] Femenino [ ] Masculino

B. Edad en años cumplidos: [ ]

C. Materias que cursa en el semestre actual: [ ]

- 
- 
1. La sintaxis de R Markdown es fácil de entender. \_\_\_\_
  2. Cuando no logro compilar un documento de R Markdown, identifico la causa y realizo los ajustes necesarios de forma inmediata. \_\_\_\_
  3. En lugar de R Markdown, prefiero utilizar hojas de cálculo y procesadores de texto para elaborar reportes de investigación o tareas escolares. \_\_\_\_
  4. Fue más difícil instalar R y los paquetes especializados, que aprender a usarlo. \_\_\_\_
  5. R Markdown hace que mis reportes de trabajo sean más fáciles de leer, entender y reproducir. \_\_\_\_
  6. Prefiero el proceso de *copiar y pegar* (gráficas, tablas y resultados) en procesadores de texto. \_\_\_\_
  7. R Markdown es muy potente para realizar cálculo estadístico, crear gráficas y tablas, fácil de aprender y redactar informes de investigación. \_\_\_\_
  8. Al inicio del curso fue difícil entender la lógica de trabajo para mezclar código en la elaboración de informes, pero ahora lo entiendo. \_\_\_\_
  9. Uso R Markdown para presentar reportes en otras materias. \_\_\_\_
- 
- 

Gracias por su atención

