

29

LA IMPUTACIÓN DE DATOS VACÍOS EN LAS MATEMÁTICAS Y SU MODELO ESTADÍSTICO

IMPUTATION OF EMPTY DATA IN MATHEMATICS AND ITS STATISTICAL MODEL

Luis German Castro-Morales¹

E-mail: ui.luiscastro@uniandes.edu.ec

ORCID: <https://orcid.org/0000-0002-7521-923X>

Wilmer Medardo Arias-Collaguazo¹

E-mail: ui.wilmerarias@uniandes.edu.ec

ORCID: <https://orcid.org/0000-0002-1438-4012>

Olga Germania Arciniegas-Paspuel¹

E-mail: ui.olgaarciniegas@uniandes.edu.ec

ORCID: <https://orcid.org/0000-0002-9161-4846>

Carlos Wilman Maldonado-Gudiño¹

E-mail: ui.carlosmaldonado@uniandes.edu.ec

ORCID: <https://orcid.org/0000-0001-8784-211X>

¹Universidad Regional Autónoma de Los Andes. Ecuador

Cita sugerida (APA, séptima edición)

Castro-Morales, L. G., Arias-Collaguazo, W. M., Arciniegas-Paspuel, O. G., & Maldonado-Gudiño, C. W. (2022). La imputación de datos vacíos en las matemáticas y su modelo estadístico. *Revista Conrado*, 18(S2), 272-280.

RESUMEN

En este artículo se analiza el uso de modelos estadístico y matemático de tipo educativo y tiene como objetivo diseñar la imputación de datos vacíos en las matemáticas y su modelo estadístico, utilizando metodología basada en la modalidad cualitativa, para el desarrollo del presente trabajo investigativo, además se utilizó el métodos teórico y empírico, adaptando técnicas como son la de documentales y la observación a través de fichas bibliográficas y fichas de observación. Los resultados más relevantes es imputar los datos faltantes que se puede adaptar a usos prácticos como el campo del estudio socioeconómico y de la bioestadística.

Palabras clave:

Modalidad cualitativa, métodos teórico y empírico, fichas bibliográficas, socioeconómico, bioestadística

ABSTRACT

This article analyzes the use of statistical and mathematical models of educational type and has as objective to design the imputation of empty data in mathematics and its statistical model, using methodology based on the qualitative modality, for the development of the present investigative work, in addition the theoretical and empirical methods were used, adapting techniques such as documentary and observation through bibliographic cards and observation cards. The most relevant results are the imputation of missing data that can be adapted to practical uses such as the field of socioeconomic studies and biostatistics.

Keywords:

Qualitative modality, theoretical and empirical methods, bibliographic records, socioeconomic, biostatistics

INTRODUCCIÓN

A medida que aumentan los volúmenes de datos, estos tienen más errores e inconsistencias que son difíciles de corregir. Hoy en día la información para uso estadístico puede ser resultado de la fusión de múltiples bases de datos y contener variables de diferente naturaleza Zúñiga, Hernández, & Jiménez (2018). Esta recopilación heterogénea de datos puede dar lugar a la falta de muchos valores: las muestras solo vienen con una fracción de las características observadas. Aunque existe una variada literatura sobre el tratamiento de los valores faltantes, por lo general se centra en la estimación de parámetros y su varianza en presencia de valores faltantes en un solo conjunto de datos Meléndez, Bolívar, & Rojano (2020). Por el contrario, hay pocos estudios de entornos de aprendizaje supervisado donde el objetivo es predecir una variable objetivo dadas las variables de entrada. En lugar de modelos generativos, estas configuraciones solo requieren modelos discriminativos o condicionales (Leyva, et al., 2018).

Por otra parte, el uso de múltiples bases de datos y fuentes de información permite que los valores faltantes pueden aparecer por una variedad de razones. Por ejemplo, cuando el encuestador, aplica dispositivos y/o aplicaciones digitales estos valores faltantes pueden surgir de una falla del dispositivo Bello, Cuta, & García (2019). Por el contrario, se pueden encontrar valores faltantes informativos en los datos de la encuesta, por ejemplo, donde los participantes pueden no responder preguntas sensibles relacionadas con opiniones impopulares (Cabrera-Álvarez & Escobar, 2019). En estudios médicos, algunas mediciones pueden ser poco prácticas en pacientes en estado crítico, en cuyo caso la presencia de valores faltantes puede estar relacionada con la variable de interés, objetivo de la predicción (Leyva, et al., 2018). Estos diversos escenarios conducen a diferentes mecanismos de valor faltante.

Al revisar sobre el tema de valores faltantes, es importante mencionar a (Rubin, 1976) quién lideró el estudio de este fenómeno estadístico, y define mecanismos de valores faltantes basados en la relación entre la falta y los valores observados: si son independientes, se dice que el mecanismo falta completamente al azar; si la falta solo depende de los valores observados, entonces es una pérdida aleatoria; de lo contrario, es una pérdida no aleatoria. Sin embargo, estos conceptos rara vez se ha discutido en el contexto del aprendizaje supervisado, teniendo en cuenta la variable objetivo de la predicción (Correa & Vélez, 2014).

Existen varios métodos estadísticos que abordan los valores faltantes como (Schafer & Olsen, 1998; Efron, 1994; Pigott, 2001; Van Ginkel et al, 2020) en América Latina tales como (Muñoz & Álvarez, 2009; Ayala & Melo, 2007). Sin embargo, la eliminación de observaciones incompletas puede permitir entrenar el modelo con datos completos. No obstante, no es suficiente para el aprendizaje supervisado, ya que el conjunto de pruebas también puede contener datos incompletos. Por lo tanto, el procedimiento de predicción debe manejar los valores faltantes.

Una solución muy común es reemplazar los valores faltantes, con valores admisibles, que sean muy cercanos para producir un conjunto de datos completo. El beneficio de la imputación es que adapta las bases de datos, así como los sistemas informáticos existentes a la presencia de valores faltantes. Sin embargo, la práctica generalizada de imputar con la media de la variable en las entradas observadas tiene serios inconvenientes, ya que distorsiona las distribuciones conjuntas y marginales de los datos, lo que induce un sesgo en los estimadores (Muñoz & Álvarez, 2009).

En este sentido, la imputación media se ha estudiado muy limitadamente, realmente su uso ha sido cuando el objetivo es predecir un resultado. Por lo tanto, la imputación en sí misma debe revisarse para configuraciones de predicción fuera de muestra: los usuarios recurren a diferentes estrategias, como imputar por separado los conjuntos de datos de prueba o imputarlos conjuntamente (Correa & Vélez, 2014). Las estrategias más elaboradas se basan en el uso de la maximización de expectativas, para ajustar un modelo a datos incompletos (Ayala & Melo, 2007).

En este trabajo, estudiamos las herramientas clásicas de valores faltantes en el contexto del aprendizaje supervisado. Basados en resultados obtenidos durante el confinamiento por el COVID 19 que ocasionó una crisis socio-educativa, donde la educación sufrió cambios radicales en su nueva modalidad de enseñanza, apartándose de la presencialidad, la educación semipresencial y la educación a distancia y profundizando en la virtualidad, una modalidad basada en la utilización de los recursos tecnológicos que ofrece la web (Crisol, Herrera, & Montes (2020); Leyva, et al., 2021), la cual exigen mucha mayor responsabilidad al estudiante volviéndose el protagonista y generador de su propio aprendizaje en esta nueva forma de educación, lo que obliga a la utilización de recursos tecnológicos los cuales no están disponibles para la mayoría de los actores del proceso educativo.

La evolución de la tecnología en nuestro entorno y el ser considerados como nativos digitales afecta considerablemente a los actores del proceso educativo en todos

los niveles de educación, obligando a incursionar en el manejo de los tics de la información y el conocimiento, razón por la cual en el año 2020 en el Ecuador hubo un crecimiento del 1,5 % de los 10.17 millones de usuarios que representa el 57 % de la población utiliza el servicio de internet es decir 147 mil nuevos usuarios y en caso de las redes sociales el 98% de los ecuatorianos ingresan a través de su dispositivo móvil. El ministerio de telecomunicaciones a través del programa Ecuador conectado tiene como meta conectar al 98 % de las parroquias existentes en el país asumiendo el desafío de asegurar el acceso de la población a los tics y poder reducir de esta manera la brecha digital existente en los diferentes sectores sociales y aprovechar los beneficios que ofrece la Sociedad de la Información y del Conocimiento (Ministerio de Telecomunicaciones y de la Sociedad de la Información, 2022)

La Matemática es una de las ciencias duras en la cual se apoyan las demás ciencias, siendo una parte de esta la Estadística la cual es una ciencia que basada en modelos matemáticos no solo se encarga de recopilar, organizar, presentar, analizar e interpretar una serie de datos numéricos, sino que permite percibir de otra manera y tener una perspectiva entre las realidades y los datos con el propósito de tomar decisiones oportunas y verosímiles en beneficio social como personal. Berenson, Levine, & Krehbiel (2006). La nueva modalidad de estudio en el siglo XXI como es la virtualidad ha ocasionado que los actores del proceso educativo profundicen en el uso de las tecnologías, de plataformas digitales tanto para impartir la clase sincrónica, como también para que realicen las actividades asincrónicas los discentes, lo cual exige responsabilidad al educando para el cumplimiento de las actividades planificadas para el periodo académico de cada una de las asignaturas, logrando evidenciar la falta de cumplimiento en estas de acuerdo a la planificación curricular como a los tiempos establecidos en las plataformas para cada una de ellas. (Leyva Vázquez, et al., 2021)

Estos inconvenientes en el cumplimiento de las tareas, de acuerdo con entrevistas realizadas a los estudiantes se debe a problemas de conectividad tanto de internet como de energía eléctrica, a que algunos estudiantes no se encuentran en los centros poblados y la cobertura de los servicios es de mala calidad, a la que su mantención para el estudio los cubre con el salario que perciben por el trabajo desarrollado en empresas públicas o privadas, lo provee la falta de tiempo para cumplir con las actividades académicas.

La consecuencia de no realizar las actividades académicas, el que no entregue a tiempo las mismas, ya sea

porque la plataforma se cerró, colapso el sistema o el internet, corte de energía eléctrica, se debe a que el estudiante no estima con responsabilidad la planificación de sus acciones cotidianas ya sea en lo laboral como el académico, generando así en el sistema de notas o calificaciones, espacios vacíos considerados como datos faltantes lo que ocasiona la disminución en su promedio, en muchos casos este valor se encuentra por debajo de las niveles que exigen los centros de educación.

Para el manejo de los datos faltantes existe dos caminos que son la eliminación de los datos y la imputación de estos, siendo este el camino más viable al cual se lo puede aplicar el valor promedio de los mismos siempre y cuando los datos faltantes no superen el diez por ciento, como también el método de la razón al comparar dos columnas de datos

MATERIALES Y MÉTODOS

Esta investigación es de carácter cuali-cuantitativo ya que permite recopilar la información, para su posterior análisis y descripción del comportamiento de los datos con el apoyo de modelos estadísticos para su interpretación y la correcta toma de decisiones en la adecuada asignación de valores de los datos faltantes. Basado en métodos científicos como el analítico y sintético que permitieron la búsqueda de información en fuentes primarias y secundarias para las bases teóricas de la investigación, además de apoyarse en métodos empíricos como la observación, la medición y el experimento que permiten la confrontación de los datos cuantitativos observados y su respectivo análisis.

Los datos faltantes en esta investigación se definen como los valores que no se encuentran disponibles en la plataforma como resultado de que el estudiante no entregó o no cumplió con las tareas enviadas debido a un sin número de factores en ciertos casos ajenos a la voluntad de los estudiantes, como consecuencia de la nueva modalidad de educación, siendo estos datos experimentales ya se tuvo el control de parte del investigador en la generación de estos, es por eso que no existe una manera satisfactoria para su manejo por lo cual se debe optimizar su recolección y concienciar el actuar de los involucrados.

La imputación de datos no es sino la forma en que se completan ciertos datos faltantes de un grupo o actividad, para que exista una consistencia y estos sean analizados de manera categórica mediante técnicas estadísticas, para el presente trabajo se utilizaron el promedio o media aritmética aplicada a cada una de las tareas desarrolladas con la fórmula , como también el modelo de la

razón entre la sumatoria de los datos de las columnas a comparar por el valor de la consigo al valor a remplazar para lo cual se aplica la fórmula

La eliminación de los datos deja a estas de manera que se puedan recuperar, mientras que el borrado de los datos es de manera permanente lo que hace ver que son dos actividades similares, pero no con la misma acción

RESULTADOS

Luego de la revisión de la plataforma y de importar las calificaciones de las actividades desarrolladas para estimar el promedio de la primera parcial de la asignatura de matemática se consolido la información que se detalla en la tabla 1.

Tabla 1. Calificaciones de las actividades desarrolladas

Alumno	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6	Tarea 7	Promedio solo considerando las tareas entregadas
1	7	9	-	9	9	10	6.3	8.4
2	9	9	9	9	9	10	7.2	8.9
3	10	10	9	8	9	9.29	7.2	8.9
4	9	7	-	8	9	7.14	5.2	7.6
5	9	10	8	6	8	10	7	8.3
6	9	9	7	9	9	10	5.4	8.3
7	10	9.5	8.5	7	9	8.57	7.2	8.5
8	9	10	9	7	9	10	8.2	8.9
9	10	9	8	8	8	7.14	2.2	7.5
10	7	10	7.5	6	8.5	8.57	8	7.9
11	-	9	7.5	8	9	10	7.2	8.5
12	10	10	8	6	8	10	8	8.6
13	10	10	8.5	8	8.5	7.14	7.4	8.5
14	7	9	9	6	9.5	10	7.8	8.3
15	10	9	8	7	9	7.14	5.5	7.9
16	10	9	8.5	9	9	8.57	7.2	8.8
17	-	10	8.5	8	9	-	8.2	8.7
18	10	10	8.5	8	8.5	8.57	8	8.8
19	7	8	9	-	8	10	-	8.4
20	10	10	9.5	10	8	10	8	9.4
21	7	9	9	7	8	8.57	6	7.8
22	8	10	8	8	9	7.14	3.6	7.7
23	7	9	7	8	9	10	7.8	8.3
24	8	10	8.5	7	8	8.57	7.2	8.2
25	7	10	-	5	9	10	2.5	7.3
26	7	10	9	7	10	8.57	6.5	8.3
27	9	10	9.5	10	8	10	8.1	9.2
28	7	10	8.5	8	9	10	8	8.6
29	7	9	8	6	8.5	8.57	7	7.7
30	7	9	-	5	9	10	6.5	7.8
31	10	10	8	10	9	8.57	8.6	9.2
32	10	10	9.5	9	9.5	10	8.1	9.4
33	7	9	-	-	8	8.57	5.4	7.6
34	7	9	8	7	8.5	10	3.2	7.5

35	10	7	6	9.5	10	7.14	7.2	8.1
36	7	9	8.5	9	8	7.14	5	7.7
37	10	9	7.5	7	9	10	8	8.6
MEDIA	8.5	9.3	8.3	7.7	8.7	9.0	6.7	
SUMA	298	345.5	266	269.5	323.5	324.97	239.9	

Fuente. plataforma Eva grado

La suma de los datos (Formula 1) está dada por la sumatoria de todos los valores de cada una de las actividades, mientras que la media es el promedio entre) la sumatoria de todos los datos dividida para el total de estos.

Formula 1

$$X' = \frac{\sum x_i}{n} \cong \frac{345.5}{37} = 9.3$$

Tabla 2. Calificaciones promedio considerando las tareas no entregadas

Alumno	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6	Tarea 7	Promedio considerando las tareas no entregadas como 0
1	7	9	0	9	9	10	6.3	7.2
4	9	7	0	8	9	7.14	5.2	6.5
11	0	9	7.5	8	9	10	7.2	7.2
17	0	10	8.5	8	9	0	8.2	6.2
19	7	8	9	0	8	10	0	6.0
25	7	10	0	5	9	10	2.5	6.2
30	7	9	0	5	9	10	6.5	6.6
33	7	9	0	0	8	8.57	5.4	5.4
Media								6.4

Si se considera las tareas no entregadas por los estudiantes (Tabla 2) con un valor de cero, se observa que el promedio de calificación baja considerablemente a un nivel poco satisfactorio, por lo cual no podemos asignar una calificación de cero a una evidencia no existente, sin considerar aspectos secundarios que hayan ocasionado el cumplimiento de esta actividad.

Tabla 3. Asignación de calificaciones por el método de la media o promedio

Alumno	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6	Tarea 7	Promedio aplicando el modelo de la media
1	7	9	8.3	9	9	10	6.3	8.4
4	9	7	8.3	8	9	7.14	5.2	7.7
11	8.5	9	7.5	8	9	10	7.2	8.5
17	8.5	10	8.5	8	9	9	8.2	8.7
19	7	8	9	7.7	8	10	6.7	8.1
25	7	10	8.3	5	9	10	2.5	7.4
30	7	9	8.3	5	9	10	6.5	7.8
33	7	9	8.3	7.7	8	8.57	5.4	7.7
Media	8.5 9.3 8.3 7.7 8.7 9.0 6.7							
X'								8.0

El promedio (Tabla 3) de cada una de las tareas en cada una de las columnas se determina con la sumatoria de todos los valores dividido para el total de los datos, el que se considera en cada tarea no entregada por los estudiantes, lo cual permite que las calificaciones se mantengan en un nivel satisfactorio.

Tabla 4. Asignación de calificaciones por el método de la razón

Alumno	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6	Tarea 7	Promedio aplicando el modelo de la razón
1	7	9	6.9	9	9	10	6.3	8.2
4	9	7	5.4	8	9	7.14	5.2	7.2
11	7.8	9	7.5	8	9	10	7.2	8.4
17	8.6	10	8.5	8	9	9	8.2	8.8
19	7	8	9	6.7	8	10	7.4	8.0
25	7	10	7.7	5	9	10	2.5	7.3
30	7	9	6.9	5	9	10	6.5	7.6
33	7	9	6.9	6.7	8	8.57	5.4	7.4
SUMA	298	345.5	266	269.5	323.5	324.97	239.9	
Media								7.9

Si se considera a la columna de la tarea 1 como a_i y la columna de la tarea 2 como b_i y se aplica la regla de la razón para el estudiante (11), (17)

$$a_i = \frac{\sum a_i}{\sum b_i} * b_i$$

Estudiante 11	$a_i = \frac{298}{345.5} * 9 \cong \frac{2682}{345.5}$ $a_i = 7.76 \cong 7.8$	Estudiante 17	$a_i = \frac{298}{345.5} * 10 \cong \frac{2980}{345.5}$ $a_i = 8.62 \cong 8.6$
---------------	---	---------------	--

Si se considera a la columna de la tarea 2 como b_i y la columna de la tarea 3 como a_i y aplicamos la regla de la razón para los estudiantes (1), (4), (25), (30), (33)

$$b_i = \frac{\sum b_i}{\sum a_i} * a_i$$

Estudiante 1	$b_i = \frac{266}{345.5} * 9 \cong \frac{2394}{345.5}$ $b_i = 6.92 \cong 6.9$	Estudiante 4	$b_i = \frac{266}{345.5} * 7 \cong \frac{1862}{345.5}$ $b_i = 5.38 \cong 5.4$
Estudiante 25	$b_i = \frac{266}{345.5} * 10 \cong \frac{2660}{345.5}$ $b_i = 7.69 \cong 7.7$	Estudiante 30	$b_i = \frac{266}{345.5} * 9 \cong \frac{2394}{345.5}$ $b_i = 6.92 \cong 6.9$
Estudiante 33	$b_i = \frac{266}{345.5} * 9 \cong \frac{2394}{345.5}$ $b_i = 6.92 \cong 6.9$		

Si se considera a la columna de la tarea 4 como a_i y la columna de la tarea 5 como b_i y aplicamos la regla de la razón para los estudiantes (19), (33)

$$a_i = \frac{\sum a_i}{\sum b_i} * b_i$$

Estudiante 19

$$a_i = \frac{269.5}{323.5} * 8 \cong \frac{2156}{323.5}$$

$$a_i = 6.66 \cong 6.7$$

Estudiante 33

$$a_i = \frac{269.5}{323.5} * 8 \cong \frac{2156}{323.5}$$

$$a_i = 6.66 \cong 6.7$$

Si se considera a la columna de la tarea 5 como a_i y la columna de la tarea 6 como b_i y aplicamos la regla de la razón para los estudiantes (17)

$$b_i = \frac{\sum b_i}{\sum a_i} * a_i$$

Estudiante 17

$$b_i = \frac{324.97}{323.5} * 9 \cong \frac{2924.73}{323.5}$$

$$b_i = 9.04 \cong 9$$

Si se considera a la columna de la tarea 6 como a_i y la columna de la tarea 7 como b_i y aplicamos la regla de la razón para los estudiantes (19)

$$b_i = \frac{\sum b_i}{\sum a_i} * a_i$$

Estudiante 19

$$b_i = \frac{239.9}{324.97} * 10 \cong \frac{2399}{323.5}$$

$$b_i = 7.38 \cong 7.4$$

DISCUSIÓN

Con la aplicación de los métodos de la imputación como es el promedio y la razón se puede observar que los datos se mantienen en un nivel moderado donde la diferencia entre aplicar las dos estrategias la variación entre los promedios de cada uno de los estudiantes esta entre un intervalo de variación de 1 al 2 %, lo que no influye considerablemente en la asignación de los datos faltantes en cada una de las actividades desarrolladas.

Los datos faltantes van a ser siempre un frecuente problema en cualquier actividad o investigación realizada, por lo cual se deben considerar ciertos principios que permitan hacer inferencias estadísticas significativas muy cercanas a la realidad, ya que esto implica consecuencias y un posible efecto causal, por lo cual se debe evidenciar en lo posible las causas del porque la no existencia de un dato, o el no cumplimiento de la actividad.

Considerando la tabla 1 se puede observar que la tarea dos y cinco los 37 estudiantes han cumplido con las tareas establecidas, en la tarea uno y cuatro corresponde el 5,4% a las tareas no entregadas, en la columna seis y siete el 2,7% de los estudiantes no ha cumplido con las tareas asignadas de la asignatura de matemáticas excepto la tarea de la columna tres donde el porcentaje de no realizar la actividad en los tiempos establecidos supera el 10% del límite que se establece para la imputación de datos faltantes.

Si se considera la media y la desviación estándar de la actividad tres los datos faltantes asignados al estudiante 25 considerando la regla empírica de la distribución normal se encuentra entre más menos una desviación, de los estudiantes 1, 30 y 33 se encuentra entre más menos dos desviaciones, y en el caso del estudiante cuatro entre mas menos tres desviaciones.

Si se toma en cuenta los promedios de la tabla dos, tres y cuatro se evidencia que el método de la razón es más consistente que el uso de la media aritmética y si estos valores los comparamos con el promedio, pero si se asigna un cero a la actividad no realizada, existe una variación de 1,5 puntos en promedio, razón por la cual se debe tener cuidado y actuar con flexibilidad tanto el docente como el estudiante, pero siempre en busca de la mejora continua, orientada al cumplimiento de los resultados de aprendizaje.

El total de tareas enviadas a los estudiantes para la actividad asincrónica de la primera parcial es de 259, de las cuales si se calcula el porcentaje de tareas incumplidas se obtiene que el 4,25% de estas no se han registrado en la plataforma, lo cual no supera el 10% permitido para la asignación de datos faltantes.

CONCLUSIONES

Los modelos matemáticos y estadísticos son principios fundamentales para cualquier proceso investigativo ya que permiten la comparación y confrontación de la información en un intervalo de tiempo o rango, estableciendo análisis, interpretaciones, conclusiones y una correcta toma de decisiones.

Los datos faltantes son producto del error u omisión de parte del investigador, pero en esta investigación son el resultado de no subir las tareas a las plataformas en los tiempos establecidos en muchos casos se debe a problemas de conectividad al internet, a la mala administración del valor tiempo y la falta de responsabilidad en asumir el rol de estudiante en esta nueva modalidad de estudios como consecuencia de la crisis sanitaria provocada por el covid 19.

El manejo de datos faltantes, siempre que estos no superen el 10% del total de estos ayuda a mantener los estándares de calidad de un producto, ayuda a mantener la autoestima de la persona, del grupo y de la sociedad, ya que permite estimar dichos valores con otros de otras variables seleccionadas ya que este problema siempre está latente en cualquier situación.

REFERENCIAS BIBLIOGRÁFICAS

Ayala, Y. & Melo, Ó. (2007). Estimación de datos faltantes en medidas repetidas con respuesta binaria. *Revista Colombiana de Estadística*, 30(2), 265-285.

Bello, A. M., Cuta, J. A., & García, E. K. (2019). Técnicas de imputación para datos de precipitación máxima mensual en la zona central de Boyacá. *Ingeniería, Investigación y Desarrollo*, 19(1), 64-79.

Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2006). *Estadística para administración*. Pearson Educación.

Cabrera-Álvarez, P., & Escobar, M. (2019). El efecto de la ponderación y la imputación en el sesgo de los estudios electorales en España. *Reis: Revista Española de Investigaciones Sociológicas*, (165), 45-64.

Correa, J. & Vélez, J. (2014). Una nota de cuidado sobre el efecto de datos parcialmente faltantes en la prueba de independencia 2. *Comunicaciones en Estadística*, 7(2), 191-201.

Crisol, E., Herrera, B., & Montes, R. (2020). Educación virtual para todos: una revisión sistemática. *Education in the knowledge society: EKS. Education in the Knowledge Society*, 21(15), 1-15.

Efron, B. (1994). Missing Data, Imputation, and the Bootstrap. *Journal of the American Statistical Association*, 89(426), 463-475.

Leyva, M., González, N., Hechavarría, J., Rivero, Y., & Daher, J. (2018). El diagnóstico de enfermedades desde el Análisis Inteligente de los Datos. *Revista Espacios*, 39(28), 16-22. <http://www.revistaespacios.com/a18v39n28/a18v39n28p16.pdf>

Leyva Vázquez, M. Y., Estupiñán Ricardo, J., Coles Gargay, W. S. & Bajaña Bustamante, L. J. (2021). Investigación científica. Pertinencia en la educación superior del siglo XXI. *Conrado*, 17(82), 130-135.

Leyva Vázquez, M. Y., Viteri Moya, J. R., Estupiñán Ricardo, J., & Hernández Cevallos, R. E. (2021). Diagnóstico de los retos de la investigación científica postpandemia en el Ecuador. *Dilemas contemporáneos: educación, política y valores*, 9(SPE1), 1-19

Meléndez, R., Bolívar, S., & Rojano, R. (2020). Imputación de valores perdidos y detección de valores atípicos en datos funcionales: una aplicación con datos de PM10. *Revista UIS Ingenierías*, 19(2), 1-10.

Ministerio de Telecomunicaciones y de la Sociedad de la Información (2022). *Ecuador conectado*. Ministerio de Telecomunicaciones y de la Sociedad de la Información. <https://www.telecomunicaciones.gob.ec/ecuador-conectado/>

Muñoz, J. F. & Álvarez, E. (2009). Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 7(1), 3-30.

Pigott, T. (2001). A Review of Methods for Missing Data. *Educational Research and Evaluation*, 7(4), 353-383.

- Rubin, D. (1976). Inference and Missing Data. *Biometrika*, **63(3)**, 581-592.
- Schafer, J. L. & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate behavioral research*. **33(4)**, 545-571.
- Van Ginkel, J., Linting, M., Rippe, R., & Van Der Voor, A. (2020). Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data. *Journal of Personality Assessment*, **102(3)**, 297-308.
- Zúñiga, C., Hernández, M., & Jiménez, J. (2018). Comparación de técnicas de imputación para tratar respuestas censuradas en un diseño de experimentos bivariado. *Nova Scientia*, **10(20)**, 190-191.