ENSAYOS 2019.
ANÁLISIS MULTIVARIANTE CON ENFOQUE DEPENDIENTE
EN LAS CIENCIAS DE LA ADMINISTRACIÓN COMO BASE PARA LA INNOVACIÓN

TOMO III

Juan Mejía Trejo
Coordinador

ENSAYOS 2019.
ANÁLISIS MULTIVARIANTE CON ENFOQUE DEPENDIENTE
EN LAS CIENCIAS DE LA ADMINISTRACIÓN COMO BASE PARA LA INNOVACIÓN

Juan Mejía Trejo
Coordinador

CUCEA

ENSAYOS 2019. ANÁLISIS MULTIVARIANTE CON ENFOQUE DEPENDIENTE EN LAS CIENCIAS DE LA ADMINISTRACIÓN COMO BASE PARA LA INNOVACIÓN

Juan Mejía Trejo
Coordinador

"Esta obra fue sometida a un proceso de dictamen por pares de acuerdo con las normas establecidas por el comité editorial del Centro Universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara"

Diseño de portada y editorial
Javier Salazar Acosta
Abraham Romero Torres
por Prometeo Editores

Primera edición, mes año

Centro Universitario de Ciencias Económico Administrativas
Instituto de Investigación en Políticas Públicas y Gobierno
Periférico Norte, No. 799, edificio B 202
C.P. 45100, Zapopan, Jalisco

Prometeo Editores S.A. de C.V.
C. Libertad 1457, Col. Americana
C.P. 44160, Guadalajara, Jalisco

ISBN: 978-607-98782-6-9

Impreso y hecho en México
Printed and made in Mexico

CONTENIDO

INTRODUCCIÓN

La presente obra, *Ensayos 2019. Análisis Multivariante con Enfoque Dependiente en las Ciencias de la Administración como base para la Innovación,* pretende reunir una serie de ensayos elaborados por los estudiantes del Doctorado de Ciencias de la Administración (DCA) del Centro Universitario de Ciencias Económico Administrativas (CUCEA) de la Universidad de Guadalajara (UdeG), basados en lo aprendido en la asignatura de Investigación Cuantitativa I. Dichos ensayos, se orientan en principio a realizar un ejercicio de disertación que refuerce ya sea la argumentación de su tesis en la parte metodológica o bien, sea una contribución a la materia. Para ambos casos se resalta la pertinencia de su redacción a partir de la introducción para desarrollar los conceptos y/o modelos que justifican la base de los puntos antagónicos a tratar siendo la base para realizar la discusión que permite aclarar la contribución esperada. Finalmente, se exponen los puntos de conclusión esenciales que sirvan al lector y al expositor, para estudios posteriores.

Es así, que esta obra se desglosa en diez ensayos, donde la primera obra: *Cálculo del poder estadístico: ¿Por qué no es habitual su cálculo en investigaciones científicas?* hace una abordaje de los conceptos de potencia estadística y su uso histórico en las investigaciones. Su principal aportación es el descubrimiento de escuelas que lo han tratado y los alcances de su consideración.

La segunda: *Obsolescencia de métodos cuantitativos: ¿Las investigaciones del pasado pierden su validez ante la introducción de métodos más refinados?*hace una reflexión de los métodos que permiten realizar pruebas más refinadas, lo cual revela posibilidades de insuficiencia en la demostración de la obsolescencia y validez de un modelo. Cierra de manera interesante en la conveniencia de considerar la aplicación futura de los métodos de refinación.

La tercer aportación: *Consideraciones en las metodologías cuantitativas para ciencias económico-administrativas con uso de regresión lineal múltiple* realiza una recopilacion de los saberes de la aplicación de la regresión lineal y su utilidad en las ciencias de la administración, dejando entrever más pros que contras siendo su uso una garantía de comprobación de variables en la comprobación de modelos.

En el cuarto apartado: *La Lógica Difusa, Regresión Múltiple, Red Neural Artificial para su uso en las Ciencias de la Administración.* Se presentan las principales técnicas que permiten realizar proyecciones en el tratamiento de los modelos, sus ventajas y desventajas. Concluye con una descripción de los alcances de cada técnica a fin de seleccionar la más conveniente en una investigación en las ciencias de la administración.

En el quinto ensayo: *Análisis de Propensión en Ciencias Sociales* se tiene una opción de análisis de datos que está incursionando de manera preliminar en las ciencias de la administración dado su aplicación inicial en las ciencias de la salud y las sociales. Cierra con un interesante cuadro de pros y contras de dicha técnica y la reflexión de su uso en las ciencias de la administración.

La sexta obra: *Regresión lineal simple, una técnica vigente para la obtención de resultados en investigaciones cuantitativas*, permite visualizar la condición de el uso de esta técnica estadística tan relevante a partir de un análisis bibliométrico. Concluye con los puntos de vista de la autora sobre ventajas y desventajas de la técnica con el fin de considerarlas en estudios posteriores.

La séptima contribución: *Especificidades, Limitaciones y Particularidades de la Regresión Logística en las Ciencias de la Administración*, se refiere a un comparativo de los usos más extendidos de ambas técnicas concluyendo en recomendaciones para su aplicación en la investigación de las ciencias de la administración.

El octavo ensayo*: Enfoques y validez del analisis bibliométrico como herramienta de regresión lineal simple para mostrar la relación*

entre turismo y las ciencias administrativas, hace un despliegue de la habilidad del autor por encontrar la relación de uso de la técnica de regresión con el turismo y las ciencias de la admninistración a partir de técnicas bibliométricas tradicionales y las basadas en los indicadores de Leiden. Aporta mapas de concentración por país y palabras clave.

La novena obra: *El modelo estadístico de Regresión en conceptos claros, para su difusión y aplicación en el área de administración*, realiza una contribución de los principales conceptos que sostienen a la técnica y recomendaciones de uso.

Finalmente, la obra: *El análisis de datos usando entre regresión logística y/o regresión lineal en estudios de discapacidad* el cual realiza un análisis de ambas técnicas y sus aplicaciones en temas de inclusión reportados a nivel país y la descripción de diversos trabajos relacionados.

Es deseo de la coordinación del presente trabajo, que este contribuya al ánimo del lector por conocer los proyectos que se desarrollan e informar de las oportunidades que se muestran, con el fin de dar seguimiento a la evolución de los mismos en su estancia en el posgrado.

Dr. Juan Mejía Trejo
Coordinador del DCA CUCEA
Universidad de Guadalajara

CÁLCULO DEL PODER ESTADÍSTICO: ¿POR QUÉ NO ES HABITUAL SU CÁLCULO EN INVESTIGACIONES CIENTÍFICAS?

Jorge Santiago Rodríguez García
Dr. Alejandro Campos Sánchez

Palabras clave: Potencia estadística, Error de tipo I, Error de tipo II.

INTRODUCCIÓN

Actualmente, las investigaciones en ciencias sociales y administrativas basan su diseño metodológico en el desarrollo de pruebas empíricas para contrastar los modelos teóricos propuestos frente a la realidad del fenómeno de estudio. Así pues, se aplican métodos de estadística inferencial para estudiar y analizar causas y efectos de fenómenos complejos que, para su entendimiento, se necesita el ejercicio de metodologías multivariantes para la comprobación de hipótesis.

En la actualidad, buena parte de las revistas científicas, por lo menos en ciencias sociales, publican artículos que no informan del tamaño del efecto y omiten sistemáticamente los cálculos del tamaño de la muestra y la potencia estadística del diseño (Bezeau y Graves, 2001; Crosby et al., 2008; Fidler, 2002; García, Ortega y De la Fuente, 2008; Kirk, 1996; Vacha-Haaze y Ness, 1999; Vacha-Haaze y Thompson, 1998). Estas omisiones ponen a discusión la veracidad de los hallazgos que puedan derivarse de dichos estudios y representan, en el decir de algunos de los más destacados especialistas, una de las mayores muestras de ignorancia colectiva (Cohen, 1988).

De acuerdo con Cárdenas y Arancibia (2014), entre las recomendaciones que ya hace tiempo ha realizado la *American Psychological Association* (APA), se encuentran:

a. La utilización, como práctica habitual, de los intervalos de confianza (IC; límites probables entre los que se encuentra la verdadera diferencia entre dos medias).

b. La exposición de los valores de las medias y desviaciones típicas (DT) de cada grupo.

c. La entrega de los valores exactos de probabilidad (y no los tradicionales $p<0.05$ o $p<0.01$).

d. Informar la potencia estadística de la prueba o diseño utilizado.

e. Realizar el cálculo complementario del tamaño del efecto que cuantifica la magnitud de la diferencia entre dos medias (Wilkinson, 1999).

Lo cierto es, que buena parte de estos preceptos no son considerados (Cárdenas y Arancibia, 2014).

Es por lo anterior que el objetivo del presente ensayo es, en principio, conceptualizar el término "potencia estadística", para posteriormente identificar las razones por las cuales el cálculo de este índice estadístico no es considerado en la presentación de resultados de las investigaciones expuestas en diversas publicaciones científicas, con la finalidad de aportar al debate sobre la importancia de este elemento de certeza estadística así como demostrar que la omisión de este cálculo no necesariamente se relaciona con razones éticas, sino que existen razones históricas y prácticas en el área de la investigación científica que no consideran imperativo el presentar el cálculo de la potencia como señal inequívoca de la veracidad de los resultados obtenidos.

Para conseguir el cumplimiento del objetivo propuesto, el presente ensayo se divide en tres partes. En la primera parte se desarrollará conceptualmente qué es potencia estadística y cuáles son los factores que influyen en el poder estadístico de un estudio. En la segunda parte se discutirán las razones por las cuales, desde el punto de vista de diversos autores, el análisis de la potencia estadística no es considerado en el diseño de una

investigación así como en la presentación de resultados obtenidos, partiendo e razonamientos históricos así como éticos y prácticos. Finalmente se presentan las conclusiones obtenidas mediante el análisis crítico de las fuentes consultadas.

DESARROLLO

A continuación se desarrolla, en principio, una explicación teórico-conceptual para definir "potencia estadística", para posteriormente analizar cuáles son los elementos que hay que considerar para su apropiada medición y utilización.

¿Qué es potencia estadística?

Todas las técnicas multivariantes, se basan en la relación entre variables de una muestra escogida aleatoriamente de una población; si se estuviese realizando un censo de toda una población, la inferencia estadística no es necesaria, porque cualquier diferencia o relación, por pequeña que sea, es "verdad" y existe (Hair et al., 1999). Sin embargo, rara vez un investigador realiza censos, por lo que se ve obligado a deducir inferencias sobre una muestra.

Hair et al. (1999) señalan que, para realizar inferencias estadísticas, el investigador debe especificar los niveles aceptables del error estadístico, y que el modo de aproximación más común es ponderar el nivel de error de tipo I, conocido como alfa (α); este tipo de error es la probabilidad de rechazar la hipótesis nula cuando es cierta, es decir, la posibilidad de que la prueba muestre significación estadística cuando en realidad no la tiene (un "positivo falso"). Asimismo, el autor señala que una vez determinado el nivel de error de tipo I, el investigador debe determinar un error asociado, llamado error de tipo II o β (beta), el cual supone la probabilidad de fallar en rechazar la hipótesis cuando es realmente falsa (un "falso negativo").

Relacionado con lo anterior, y de acuerdo con Reyes (2013), el poder de una prueba estadística radica en la probabilidad de que la prueba va a rechazar la hipótesis nula cuando la hipótesis nula

es falsa. En otras palabras, se trata de la probabilidad de tomar una decisión con base en falsos negativos (Cohen, 1988).

En palabras de Reyes (2013) "la probabilidad de que ocurra un error de tipo II, se refiere a la tasa de falsos negativos (β); donde la potencia es igual a 1- β, siendo β (beta) el error señalado" (p. 24).

Por lo tanto, representa la capacidad de una prueba estadística para detectar como estadísticamente significativas las magnitudes de diferencias o asociaciones determinadas (Díaz y Fernández, 2003). Cuando el poder estadístico aumenta, la probabilidad de cometer un error de tipo II disminuyen.

El análisis adecuado del poder estadístico de una investigación es un paso fundamental, tanto en la fase de diseño como en la interpretación y discusión de sus resultados, dado que se trata de cuantificar la capacidad que tiene el estudio para encontrar diferencias, si las hubiere (Reyes, 2013).

La probabilidad de los diferentes tipos de error en las pruebas de hipótesis se muestra en la siguiente tabla:

	No rechazo de la hipótesis nula	Rechazo de la hipótesis nula
Hipótesis nula verdadera	No hay error Probabilidad=1-α	Error de tipo I Probabiliad=α
Hipótesis nula falsa	Error tipo II Probabilidad= β	No hay error Probabilidad= 1-β

Fuente: Spybrook (2011).

Factores que influyen en el poder estadístico de un estudio.

Cárdenas y Arancibia (2014) afirman que la potencia estadística se calcula de acuerdo a 3 elementos: tamaño de la muestra (n), nivel de error (α) y tamaño del efecto. Los autores afirman que, cuanto mayor sea la muestra, mayor será la potencia estadística (manteniendo constante el tamaño del efecto y α), dado que el error aleatorio de medida es menor (Lispey, 1990; Cohen, 1988).

El tamaño del efecto representa el grado en que la hipótesis nula es falsa; cuando el tamaño del efecto es grande, la potencia estadística aumenta (Cohen, 1992). Al incrementar el error de tipo I la potencia también aumenta, y cuanto más pequeño es el valor de α más baja será la potencia (Cárdenas y Arancibia, 2014). Es por todo lo anterior que Sedlmeier y Gigerenzer (1989) señalan que debe equilibrarse la posibilidad de cometer errores de tipo I y II.

Reyes (2013) argumenta que los factores que influyen en el análisis de poder estadístico de una prueba son los siguientes:

1. **El tamaño de la muestra usada.** Determina el error de muestreo inherente al resultado de la prueba; es complicado detectar un efecto en muestras pequeñas, por lo que aumentando la muestra, se puede obtener un poder estadístico más alto.

2. **La magnitud del efecto de interés en la población.** Puede ser cuantificada en términos del tamaño del efecto; de tal manera que donde hay un poder mayor, hay un efecto mayor.

3. **El nivel de significancia estadística utilizado en la prueba.** Un nivel de significancia estadística señala lo improbable que puede ser un resultado. Reyes (2013) explica que la significancia estadística es cuánto se está dispuesto a tomar el riesgo de asumir una conclusión equivocada, para lo cual, expone que, los criterios más utilizados son las probabilidades de 0.05 (5%, 1 en 20), 0.01 (1%, 1 en 100) y 0,001 (0.1%, 1 en 1,000).

4. **La variabilidad de la respuesta o desviación estándar del estudio.** Cuanto mayor sea la variabilidad en la respuesta, más difícil será detectar diferencias entre los grupos que se comparan y menor será el poder estadístico del estudio.

En este sentido, en el diseño de la investigación debiera considerarse el tamaño de la muestra y la potencia estadística que se lograría con ella; sin embargo, en aquellos estudios donde dicho paso no se ha considerado, resulta importante exigir el cálculo y especificación del efecto, por lo menos (Cárdenas y Arancibia, 2014), lo que permitiría comprender de mejor manera los resultados de dichos análisis.

DISCUSIÓN

El concepto de "potencia estadística" se atribuye a Neyman y Pearson (1928). Fue hasta 1962 cuando apareció, en el ámbito de las ciencias sociales, un estudio sistemático de la potencia estadística gracias al trabajo de Cohen (1962) en el que se destacó, en principio, la importancia de la potencia estadística dentro de la investigación experimental proporcionando una serie de pautas para llevar a cabo un análisis de potencia. En este trabajó se alienta a los investigadores a prestar mayor atención a la potencia estadística de las pruebas, evitando centrarse únicamente en el análisis de significación (Bono y Arau, 1995).

Así, el trabajo de Cohen ha inspirado diversos trabajos sobre la potencia estadística en áreas de las ciencias sociales, así como el desarrollo de diversos programas computacionales para su medición (Goldstein, 1989).

La aplicación de la medición de la potencia estadística en la investigación científica.

El desarrollo de productos científicos no ha tenido impacto en investigaciones posteriores al trabajo de Cohen (1962). Ejemplo de lo anterior es el trabajo realizado por Sedlmeier y Gigerenzer (1989), quienes en un análisis del volumen de 1984 del *Journal of Abnormal Psychology*, identificaron que de 54 artí-

culos que contenía esa publicación sólo en dos se mencionaba la potencia estadística y en ninguno de ellos se estimaba.

El escaso interés por la medición de la potencia estadística se refleja en una serie de estudios realizados analizando diversas revistas científicas, como lo son Anales de Psicología (Sánchez et al., 1992), Revista de Psicología General y Aplicada (Valera et al., 1993), Psicológica (Frías, García y Pascual, 1993) y Anuario de Psicología (Frías, García y Pascual, 1993). Además, si se considera como evidencia los manuales tradicionales de estadística, se puede observar que con frecuencia el tema de la potencia no se trata (Bono y Arau, 1995).

Se desconoce la verdadera razón por la que los investigadores ignoran el análisis de la potencia estadística, sin embargo, Chase y Tucker (1976) argumentan que existen dos escuelas en cuanto a la utilización de las pruebas estadísticas: la escuela de Fisher y la escuela de Neyman-Pearson. En este sentido, la escuela de Fisher considera las pruebas estadísticas como pruebas de significación, mientras que la escuela de Neyman-Pearson las conceptualiza como pruebas de decisión. Esto se puede traducir en que, de acuerdo con el enfoque Fisheriano se concluiría que la hipótesis nula (H_0) no es válida, con lo que se probaría la existencia del fenómeno que se está estudiando; mientras que la corriente Neyman-Pearson simplemente rechazaría la hipótesis nula en un análisis particular.

La explicación anterior, diferenciando ambas escuelas de análisis estadístico se sintetiza en la siguiente tabla:

Escuela de Fisher	Escuela Neyman-Pearson
Considera las pruebas estadísticas como pruebas de significación.	Conceptualiza pruebas estadísticas como pruebas de decisión.
Un defensor del enfoque fisheriano concluiría que la hipótesis nula (H_0) no es válida, con lo que se prueba la existencia del fenómeno que se está estudiando.	Un investigador de la tradición de Neyman-Pearson, simplemente rechazaría la H_0 en esta ocasión particular.
Fisher daba prioridad a un nivel de significación de 0.05, nunca prescribió que tal nivel debiera mantenerse fijo o que debiera establecerse antes de llevar a cabo el experimento.	Neyman-Pearson requiere que el nivel de significación se determine antes de cualquier análisis estadístico y que el investigador se adhiera a él en todas las decisiones estadísticas.
De acuerdo con Fisher, se puede afirmar que el efecto no es cero cuando se rechaza la H_0, pero no es posible concluir que sea cero cuando se acepta.	El planteamiento de Neyman-Pearson postula la existencia de una hipótesis alternativa (H_1) exacta del tamaño del efecto. Esta proposición llevó a al concepto de error de Tipo II, relacionado con el de potencia.

Fuente: Elaboración propia, con información de Bono y Arnau (1995).

Aunado a lo anterior se podría señalar que la razón por la cual los investigadores descuidan la consideración de la potencia estadística se debe a razones históricas. Los manuales educativos difundieron, en principio, el enfoque Fisheriano y posterior a la Segunda Guerra Mundial, los académicos reconocieron el impacto del enfoque de Neyman-Pearson y a sustituir la escuela anterior (Bono y Arau, 1995). El resultado de esta evolución fue una concepción híbrida con ideas antagónicas, lo que generó confusión del significado de conceptos básicos, y a una posible explicación del continuo descuido del concepto de potencia estadística (Bakan, 1966; Oakes, 1986).

No obstante lo anterior, existen académicos como Cárdenas y Arancibia (2014) que son más severos en su razonamiento sobre las razones y los efectos de un apropiado cálculo de la potencia estadística, señalando que esta omisión lleva a tomar decisiones fundadas en el desconocimiento de una parte importante de la información y que esta ausencia elude hacerse cargo de los errores de tipo II que constituyen la prueba más relevante de validez de cualquier diseño de estudio. Asimismo, los autores señalan que el que los investigadores sólo señalen la significación estadística de los análisis estadísticos, sin especificar la potencia, conduce la mayoría de las veces a generar predicciones triviales, generando conocimiento que sobrevalora los hallazgos que se traducen en resultados contradictorios, problemas que podrían haber sido resueltos si se elevara mínimamente la exigencia sobre la validez de los hallazgos.

Lo anterior se complementa con lo comentado por Altman y Bland (1995), quienes señalan que las pruebas de significación están lejos de ser un factor de certeza y constituyen un criterio pobre al momento de aceptar o rechazar los resultados de una investigación, ya que la falta de significación estadística no significa que la hipótesis nula sea verdadera ni que los efectos de los grupos sean equivalentes.

Esta confusión y desinterés sobre la consideración de la potencia de las pruebas estadísticas sólo cambiará cuando los editores de las principales publicaciones científicas exijan, en su política editorial, que los investigadores calculen y declaren la potencia de sus pruebas de significación (Sedlmeier y Gigerenzer, 1989).

CONCLUSIONES

En la actualidad, buena parte de las revistas científicas, por lo menos en ciencias sociales y ciencias administrativas, publican artículos que no informan del tamaño del efecto y omiten exhibir los cálculos del tamaño de la muestra y la potencia estadística del diseño de la investigación.

El poder de una prueba estadística significa la probabilidad de que la prueba va a rechazar la hipótesis nula cuando la hipótesis nula es falsa. Es decir, se trata de la probabilidad de tomar una decisión con base en falsos negativos.

Se podría señalar que la razón por la cual los investigadores descuidan la consideración de la potencia estadística se debe a razones históricas. Algunos autores argumentan que existen dos escuelas en cuanto a la utilización de las pruebas estadísticas, la escuela de Fisher y la escuela de Neyman-Pearson, corrientes que con el transcurrir del tiempo fueron evolucionando a la par, lo que dio como resultado una concepción combinada de ideas contrarias, generando confusión sobre el significado de conceptos estadísticos.

No obstante lo anterior, hay quienes señalan que el no informar sobre índices estadísticos básicos lleva a tomar decisiones fundadas en la omisión de una parte valiosa de los resultados obtenidos, lo que lleva a generar conocimiento que da por cierto hallazgos que carecen de fundamentos incontrovertibles, ya que la ausencia de evidencia nunca es evidencia de ausencia de efectos (Altman y Bland, 1995).

Este desinterés sobre la consideración potencia estadística sólo cambiará cuando los editores de las publicaciones, dentro en su política editorial, exijan que los investigadores calculen y declaren la potencia de sus pruebas estadísticas.

Como se hace patente en este ensayo, la discusión de la importancia del cálculo de la potencia tiene un aspecto histórico que ha transformado el conocimiento de la disciplina estadística. Un ejemplo de lo anterior es el desarrollo de las corrientes Fisheriana y Neyman-Pearson, que ven en la estadística razones instrumentales distintas, uno como prueba de significación y otro como prueba de toma de decisiones, por lo que estos conceptos dan luz acerca de los resultados que ofrecen los científicos en las ciencias sociales y en las económico administrativas.

Si se considera a la estadística como una prueba de significación, habrá que considerar el cálculo de la potencia como un índice más que aporta evidencia sobre la validez relativa de la investigación en cuestión; pero si se considera como una prueba de toma de decisiones, es un instrumento que ayuda al investigador a diseñar su proyecto científico, cuyo objetivo primordial es el aportar a la ciencia resultados absolutos sobre la investigación que realiza. Es así pues, que mientras la corriente Fisheriana no especifica el momento en el cual hay que realizar el cálculo de la potencia, la escuela Neyman-Pearson invita al científico a calcularlo desde el inicio para que se considere si vale la pena o no continuar con el diseño original, o en un caso más extremo, determinar si vale la pena o no continuar con el proyecto de investigación.

Entonces, no se puede asegurar inequívocamente que, el que un científico que no presente el cálculo de la potencia se vincula con razones éticas (como trampear o alterar resultados), sino que su formación en el ámbito estadístico tiene el enfoque de alguna de las dos corrientes de pensamiento ya presentadas; no obstante, esto no resta importancia al aporte que ofrece el cálculo de este índice para informar a la comunidad científica acerca de la validez de los resultados que un proyecto de investigación aporta.

En este orden de ideas, se puede inferir que la omisión de la presentación del cálculo de la potencia estadística no sólo se vincula con hábitos de los investigadores, sino con las políticas editoriales de las principales revistas científicas. Resulta obvio que los investigadores publican los resultados de sus investigaciones de acuerdo a las exigencias de las editoriales, y que hasta que las mencionadas publicaciones no consideren en sus políticas imperativo el cálculo de este índice, los investigadores no se verán obligados a incluirlos en sus publicaciones.

Lo anterior requiere consenso entre las editoriales para elevar el nivel de certeza de las investigaciones que publican, de tal manera que se generen normas de carácter universal. Sin

embargo, esto se antoja complicado, ya que en el ámbito de libre mercado en el que se desenvuelve la economía, la competencia entre las editoriales tiene impacto en sus políticas; por un lado, algunas revistas pueden considerar que tales exigencias pueden disminuir el ritmo con el cual el conocimiento se genera impactando de forma directa con la velocidad a la cual publican y generan recursos económicos, sin embargo, este enfoque compromete la calidad de los resultados que se obtienen y publican. Por otro lado, el endurecimiento de las normas editoriales puede generar conocimiento de mayor calidad, comprometiendo su participación en el mercado y su salud financiera.

Como se hace evidente en el presente ensayo, el impacto del debate no sólo se relaciona con la calidad de la investigación científica, sino que tiene influencia directa en las prácticas editoriales y, por ende, en cuestiones de carácter comercial y del espíritu de la generación de conocimiento. Por lo que se puede invitar a los investigadores y a los responsables de las principales editoriales a analizar y debatir el verdadero carácter de la generación de conocimiento:

1. Si el conocimiento científico es visto como una mercancía y un medio para generar recursos económicos, y lo que se necesita en esta materia es impulsar el volumen de las publicaciones, ¿se pone en riesgo la veracidad y calidad de los hallazgos?

2. Si el conocimiento es considerado como un instrumento para mejorar las condiciones de la humanidad, y por ello es necesario endurecer las políticas editoriales con la finalidad de ofrecer resultados veraces y fundamentados, ¿se pone en riesgo el modelo económico de las editoriales?

REFERENCIAS

1. Altmand, D.G., y Bland, J.M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal*, 311, 485.
2. Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 432-437.
3. Bezeau, S. y Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychology research. *Journal of Clinical and Experimental Neuropsychology*, 23(3), 399-406.
4. Bono, R., y Arnau, J. (1995). Consideraciones generales en torno a los estudios de potencia. *Anales de psicología*, 11(2), 193–202.
5. Cárdenas, M., y Arancibia, H. (2014). Potencia estadística y cálculo del tamaño del efecto en G*POWER: complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud y Sociedad*, 5(2), 210–224.
6. Chase, L.J. y Tucker, R.K. (1976). Statistical power: derivation, development, and data-analytic implications. *The Psychological Record*, 26, 473-486.
7. Cohen, J, (1988). *Statistical Power Analysis for the Behavioral Sciences*, (2ª. ed.), New Jersey: Lawrence Erlbaum Associates.
8. Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
9. Cohen, J. (1992). Cosas que he aprendido (hasta ahora). *Anales de Psicología*, 8(1-2), 3-18.
10. Crosby, R.D., Wonderlich, S.A., Mitchell, J.E., de Zwaan, M., Engel, S.G., Connolly, K., Flessner, C., Redlin, J., Markland, M., Simonich, H., Wright, T.L., Swanson, J.M., y Taheri, M. (2008). An empirical analysis of eating disorders and anxiety disorders publications (1980-2000) – part II: Statistical hypothesis testing. *International Journal of Eating Disorders*, 39(1), 49-54.
11. Diaz, P. y Fernández, P. (2003). Cálculo del poder estadístico de un estudio. Complejo Hospitalario-Universitario Juan Canalejo. La Coruña, España. *Cad Aten Primaria*, 10. 59-63.

12. Fidler, F. (2002). The Fifth edition of the APA Publication Manual: Why its Statistics Recommendations are so Controversial. *Educational and Psychological Measurement*, 62(5), 749-770.

13. Frias, D., García, J.F. y Pascual, J. (1993). Estudio de la potencia de los trabajos publicados en "Psicológica". Estimación del número de sujetos fijando α y β. *Actas del III Simposium de Metodología de las Ciencias Sociales y del Comportamiento*, Santiago de Compostela, La Coruña.

14. García, J., Ortega, E., y De la Fuente, L. (2008). Tamaño del Efecto en las revistas de Psicología Indizadas en Redalyc. *Informes Psicológicos*, 10(11), 173-188.

15. Goldstein, R. (1989). Power and sample size via MS/PCDOS computers. *American Statistician*, 43, 253-260.

16. Hair, J. F., Anderson, R., Tatham, R., y Black, W. (1999). *Análisis multivariante* (5ª. ed.). Madrid, España: Pearson Educación.

17. Kirk, R.E. (1996). Practical Significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759.

18. Lipsey, M.W. (1990). *Design sensivity: Statistical power for experimental research.* Newbury Park, CA. Sage.

19. Neyman, J. y Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20A, 175-240.

20. Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.

21. Reyes, M. (2013). El poder estadístico: diferencias observadas cuando se cambia el alfa establecido en un estudio de investigación. *Scientific International Journal*, 10(1), 23–32.

22. Sánchez, J., Valera, A., Velandrino, A.P. y Marín, F. (1992). Un estudio de la potencia estadística en Anales de Psicología (1984-1991). *Anales de Psicología*, 8, 19-32.

23. Sedlmeier, P. y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies?. *Psychological Bulletin*, 105, 309-316.

24. Sedlmeier, P., y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies?. *Psychological Bulletin*, 105(2), 309-316.

25. Spybrook, J. and Bloom, H. et.al. (2011). *Optimal Design Plus Empirical Evidence*. William T. Grant Foundation.

26. Vacha-Haase, T. y Ness, C.M. (1999). Statistical significance testing as it relates to practice: Use within Professional Psychology: Research and Practice. *Professional Psychology: Research and Practice*, 30(1), 104-105.

27. Vacha-Haase, T., y Thompson, B. (1998). Further Comments on Statistical Significance Tests. *Measurement and Evaluation in Counseling and Development*, 31(1), 63-67.

28. Valera, A., Sánchez, J., Velandrino, A.P. y Marín, F. (1993). Un estudio de la potencia estadística de la revista de Psicología General y Aplicada (1990-1992). *Actas del III Simposium de Metodología de las Ciencias Sociales y del Comportamiento,* Santiago de Compostela, La Coruña.

29. Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.

OBSOLESCENCIA DE MÉTODOS CUANTITATIVOS: ¿LAS INVESTIGACIONES DEL PASADO PIERDEN SU VALIDEZ ANTE LA INTRODUCCIÓN DE MÉTODOS MÁS REFINADOS?

Elías Alejandro García Gutiérrez
Dr. Juan Antonio Vargas Barraza

Palabras clave: Métodos cuantitativos, obsolescencia, validez

INTRODUCCIÓN

En el año 2012 IBM presentó la versión 21 del *Statistical Package for the Social Sciences* (SPSS), uno de los programas de análisis de datos más utilizados tanto por la academia como por la industria. En dicha actualización se introdujo la corrección de Lilliefors para las pruebas de normalidad de Kolmogorov-Smirnov (IBM, 2018). La prueba de Lilliefors, cuando la media y varianza de la población son desconocidas, propone un método más conservador (Lilliefors, 1967), o estricto, para determinar la normalidad en un grupo de datos, que la prueba clásica de Kolmogorov-Smirnov (K-S), propuesta por Nikolai Smirnov en 1939.

Gran variedad de métodos cuantitativos de análisis operan bajo el supuesto de normalidad de sus datos, por lo que si sus datos no respetan el supuesto de normalidad, el método pierde validez (Zimmerman, 1998; Van Horn, Smotht y Fagan, 2012), y por consiguiente, pierde validez la investigación que empleó dichos métodos.

Lo anterior es sólo un caso en el que una prueba reconocida por aportar validez a los datos de investigación, en este caso la K-S, se vuelve cuestionada por la introducción de una prueba más refinada, la de Lilliefors, al punto de que la nueva prueba se convierte en un estándar para la academia, en este caso representado por su inclusión en el SPSS, volviéndose la prueba anterior insuficiente para demostrar validez. Este ejemplo de evolución o

refinamiento de los métodos cuantitativos de análisis, probablemente uno de muchos, se presta para elaborar interesantes cuestionamientos a los conceptos de validez y obsolescencia:

¿Qué ocurre en un mundo académico pos-Lilliefors con las investigaciones que utilizaron datos validados bajo la prueba K-S, cuando la academia acordaba que esa prueba demostraba validez, y que no pasarían la más estricta prueba de Lilliefors? ¿Todas esas investigaciones pierden validez? Hay que considerar que la prueba de Lilliefors se propuso en 1967, veintiocho años después de la de Smirnov, y además, se agregó al SPSS hasta su versión 21, después de veinte versiones del programa y 73 años después de que Smirnov publicara "Estimados de desviación entre funciones de distribución empírica en muestras independientes". ¿Cuántas investigaciones existen que se validaron con K-S y que potencialmente no estarían validadas bajo Lilliefors? Y ese es un supuesto entre todos los supuestos estadísticos para el análisis de datos, siendo el análisis de datos sólo un paso entre todos los pasos para la investigación cuantitativa. ¿Cuántos métodos, procesos, pruebas, algoritmos y elementos no han cambiado con el tiempo, redefiniendo su concepto de validez? ¿Qué se debería hacer con toda esa información potencialmente obsoleta? ¿Revalidar investigación por investigación, aceptarla con los estándares que existían en el tiempo que se publicó o dejarla en el pasado y suponer que nunca existió?

Esta disertación buscará dar respuesta a esas preguntas mediante el análisis de los conceptos de obsolescencia y validez, lo que permitirá al investigador tomar conclusiones acerca de la certeza y legitimidad tanto de los datos citados como los generados en investigaciones propias.

A continuación se analizarán los conceptos de obsolescencia y validez, se expondrán métodos para maximizar la validez y minimizar la obsolescencia, y se propondrán conceptos capaces de suplir a la validez desde una perspectiva epistemológica, para concluir dando respuesta a las preguntas planteadas.

Obsolescencia

El estudio de la obsolescencia comenzó al comparar el desuso de las investigaciones académicas con el tiempo que tardan las sustancias radioactivas en perder sus efectos por mitad, "half-life" (Arao, Veloso y da Silveira, 2015), esta obsolescencia tiene la forma de una función exponencial, varía entre disciplinas o ramas de la ciencia, y se puede acelerar por la poca calidad en la información o cambios en la técnica (Burton y Kebler, 1960), de hecho, las ciencias duras que dependen más de la tecnología, son las que presentan una obsolescencia especialmente rápida (Cunningham y Bocock, 1995). Las ciencias sociales, incluyendo las económicas y administrativas, sufren de menor obsolescencia, sin embargo, las humanidades, como historia y filosofía, las cuales se fundamentan en literatura archivada, resultan las más duraderas (Song, Ma y Yang, 2014).

El concepto de la obsolescencia ha recibido poca atención a través de los años, los investigadores han dado el tema como probado a pesar de que se trata tan solo de una hipótesis (Line, 1993), hipótesis, que a pesar de que muestra una serie de factores multivariantes en forma de tendencias, se ha degradado en una supuesta norma que estipula que las investigaciones se invalidan con el tiempo, sin considerar los factores individuales de cada una de ellas, sencillamente por la constante generación de investigaciones nuevas (Russell, 2010).

Esta invalidación temporal automática resulta tan incongruente, que los mismos Burton y Kebler destacan la existencia de "literatura clásica", literatura que sin importar el paso del tiempo no pierde relevancia (1960). Así en las ciencias administrativas tenemos a autores como Mayo y Drucker, quienes a través de los años continúan siendo citados; en la filosofía se sigue estudiando a Sócrates, Platón y Aristóteles, y hasta en las matemáticas, la más dura de las ciencias duras, Pitágoras, quien completara toda su obra en 475 antes de Cristo, continúa siendo un referente.

En otro contraargumento ante la obsolescencia temporal, Oberhofer destaca que los cambios en uso de un documento

no necesariamente están relacionados con su validez y utilidad, que la disminución en el uso de un documento puede ocurrir a pesar de que el mismo continúe siendo internamente válido y potencialmente útil (1992). De hecho, hay casos en los que una teoría cae en desuso, para años después volver a capturar el interés de la academia. En otras palabras, y a pesar de que para efectos prácticos se utilicen de manera indistinta, la popularidad académica (número de citas, interés académico por el tema) y la validez no son lo mismo: las investigaciones académicas mantienen o pierden su validez por méritos propios, por la cimentación o descubrimiento de nuevos hallazgos dentro de las investigaciones mismas y los métodos que las sustentan, no por el paso natural del tiempo, no por la atención que les presten otros investigadores, y no por una agrupación que pueda ser estudiada mediante tendencias.

Validez

La validez es la característica que indica si los resultados de una investigación son correctos, repetibles y generalizables, si resultan isomórficos a la realidad (Kirchgässler, 1991). Las investigaciones científicas basan en el concepto de validez la generación de conocimiento confiable, y ésto resulta especialmente cierto para las ciencias duras, las cuales utilizan métodos cuantitativos de análisis de datos por la naturaleza generalizable de los números.

Sin embargo, algunos académicos consideran que el concepto de validez debería ser abandonado en las investigaciones cualitativas, y en especial, en las ciencias sociales y humanidades, las cuales tratan con elementos humanos y no con fórmulas y números (Avis, 2006; Hayashi, 2019). Existen investigaciones varias que al relegar el concepto de validez a los métodos cuantitativos, desarrollan conceptos equivalentes para los métodos cualitativos, como "certeza", (Weber, 1922/1968) y "veracidad" (Koch, 1994/2006). Para poder integrar estas nociones dispares entre los métodos cuantitativos y cualitativos para su uso en métodos mixtos, Onwuegbuzie y Johnson proponen combinar el concepto cuantitativo de validez con el cualitativo de vera-

cidad en la "legitimidad", un parámetro integrado que consiga inferencias creíbles, confiables, veraces, transferibles y/o confirmables (2006).

Validez interna y externa

Demostrar la forma en la que el conocimiento adquirido se apega a la realidad resulta complejo, por lo que el concepto de validez se suele dividir de diversas maneras, y la más popular entre ellas es "validez interna y validez externa": Campbell define la validez interna como la certeza de que el estímulo experimental ocasione una diferencia significativa en el escenario específico observado, la certeza de que el experimento esté probando lo que se está buscando que pruebe; mientras que la validez interna es la representatividad, la generabilidad, las poblaciones, escenarios y variables en los que los efectos encontrados pueden ser generalizables (1957).

A pesar de que los conceptos de validez interna y externa son ampliamente aceptados y puestos en práctica, la academia ha encontrado limitaciones con la validez externa: para que un modelo cuantitativo sea un espejo de la realidad debe considerar las variables y factores potencialmente infinitos que afectan a la realidad, lo cual resulta imposible, por lo tanto, es preferible enfocarse en la validez interna, a pesar de que la investigación quede limitada a una pequeña fracción de la realidad percibida (Calder, Phillips y Tybout, 1982). Inclusive el mismo Campbell argumenta que si una validez tiene que sacrificarse por el beneficio de la otra, se le debe dar la prioridad a la validez interna (1957).

La conclusión lógica de lo expuesto anteriormente es que a pesar de que los métodos cuantitativos son considerados como más replicables y generalizables que los métodos cualitativos, contienen limitaciones conocidas en cuanto a su capacidad de generalizar y reflejar la realidad de forma holística, por lo que se le da prioridad a mantener la consistencia interna del estudio, a realmente corroborar su universalidad. Sin embargo, en estas limitaciones y priorizaciones se observa una paradoja, ya que

la validez interna se obtiene mediante métodos cuantitativos de análisis y valoración de datos, los cuales se utilizan de forma general por su supuesta replicabilidad, universalidad y validez externa, la cual los mismos académicos declaran limitada y defectuosa, uno de sus defectos siendo precisamente el ejemplo de la prueba K-S contra Lilliefors que motivó esta disertación. Si la validez externa está limitada, la validez interna que depende de procesos con validación externa que la validen, sufrirá de las mismas limitaciones.

Métodos para maximizar la validez y minimizar la obsolescencia

Para mejorar la validez y disminuir la obsolescencia algunos académicos proponen el uso de las técnicas más avanzadas de análisis y la implementación de tecnología (Van Aken, 2005), sin embargo, a pesar de que ésto aumentaría la percepción de validez mientras dichas técnicas mantengan la percepción de actualidad, la historia nos muestra que en algún punto futuro esas técnicas serán consideradas obsoletas, a la vez que las investigaciones que las emplearon.

Otra opción más conservadora es el utilizar múltiples métodos de análisis (Meijer, Verloop, Beijaard, 2002), estos criterios pueden inclusive combinar métodos cualitativos y cuantitativos para minimizar lo más posible la obsolescencia del sistema (Adetunji, Bischoff y Wully, 2017), de esta forma, la investigación tendrá la vida útil del último método que se considere relevante de todos los que se utilizaron para evaluar.

Sin embargo, estas técnicas multimetódicas lo único que consiguen es alargar la vida útil de la investigación, no eliminan la obsolescencia ni aseguran el mantenimiento de la validez. Esto se debe a que la validez, como criterio de evaluación de trabajos académicos y científicos, tiene un problema desde niveles epistemológicos.

La epistemología de la validez

El buscar una definición epistemológica de validez sugiere nuevas complicaciones, ya que al considerar que diferentes investigaciones utilizan diferentes perspectivas epistemológicas se podría concluir que cada una de esas investigaciones podría tener diferentes definiciones para el concepto de validez. Sin embargo, a pesar de que no es posible seleccionar alguna corriente epistemológica por sobre otra, sí resulta posible descartar algunas, que de acuerdo a los usos y acuerdos de la academia, no tendrían cabida dentro del método científico.

El dogmatismo, del griego "doctrina fijada", es la corriente epistemológica pre-socrática para la cual no se encuentra en discusión el tema del conocimiento, ya que da por supuesto el contacto directo entre el sujeto y el objeto; el sujeto, la conciencia cognoscente, se apropia del objeto, el conocimiento (Hessen, 1926). El principal ejemplo del dogmatismo son las religiones, en las que por medios metafísicos se le revela a la humanidad la verdad innegable e incuestionable. Si la investigación académica pudiera ser regida por el dogmatismo, la verdad ya se poseería, no tendría que buscarse, y por lo tanto, no habría necesidad o cabida para la investigación.

En el polo opuesto se encuentra el escepticismo, que niega por completo el contacto entre sujeto y objeto, imposibilita al sujeto a hacer cualquier juicio acerca del objeto de estudio (Hessen, 1926). Si la academia buscara regirse por el escepticismo, tendría que aceptar que le es imposible generar conocimiento, y por lo tanto, cualquier investigación resultaría inútil.

Resulta claro que la investigación académica no se puede alinear en los extremos del dogmatismo o escepticismo, no puede ni asegurar que poseé el conocimiento absoluto ni negar la posibilidad de obtenerlo, o de lo contrario anularía el valor de la investigación misma. Y sin embargo, cuando se habla de que una investigación poseé o carece de validez, pareciera que se toman precisamente posiciones dogmáticas o escépticas: o se valida o se rechaza la hipótesis, o se valida o se rechaza la investigación,

o se acepta o se niega que la investigación representa la realidad; rara vez se considera que la hipótesis, que la investigación, pudiera únicamente aproximarse a la verdad, pero sin llegar a alcanzarla, sin llegar a explicarla.

Quizá esta incoherencia epistemológica de los métodos cuantitativos, que valora la validez por sobre los métodos cualitativos, se deba a su uso dogmático de las matemáticas "perfectas": "si se demuestra matemáticamente debe ser cierto".

No es objetivo de esta disertación el demostrar o contradecir la exactitud de los sistemas matemáticos, sin embargo, resulta claro que si los sistemas matemáticos son perfectos, los académicos aún no llegan a comprenderlos del todo, y la prueba está en que los métodos estadísticos y las comprobaciones matemáticas no paran de refinarse, de evolucionar, mientras que lo perfecto, lo exacto, no requiere cambiar, ya es perfecto tal como es. Por lo tanto, si el conocimiento es validado dogmáticamente por la perfección de sistemas matemáticos, se tendría que aceptar que tanto el conocimiento del pasado, como el conocimiento actual, no son válidos, habría que esperar a que los científicos se hagan de esos sistemas matemáticos perfectos en su totalidad, para entonces sí poder validar el conocimiento generado hasta ese momento.

Legitimidad, utilidad y popularidad

Si una validez demostrada matemáticamente no es capaz de asegurar que una teoría refleja la realidad, si escuelas cualitativas niegan la utilidad del concepto de validez, y si en la academia en general, la validez llega a pasar a segundo plano al momento de mantener viva una obra, de evitar que caiga en la obsolescencia, quizá sería conveniente abandonar tal concepto dicotómico e inexacto para remplazarlo por un concepto, o una serie de conceptos, que sugieran sus verdaderas características y aplicación. Los conceptos que se proponen son: legitimidad, utilidad y popularidad.

Legitimidad

La legitimidad no necesita lidiar con los cuestionamientos de la epistemología y gnoseología de la verdad y el conocimiento, ya que no asegura que los resultados de una investigación sean verídicos y se apeguen a la realidad, únicamente evalúa si los resultados de una investigación son considerados por la comunidad científica y académica como reales y veraces, por apegarse al rigor de los métodos de generalización y réplica acordados por ésta.

La legitimidad puede perderse únicamente después de que una teoría que contradiga a la anterior se legitime, no por el tiempo, no porque una sola de sus partes o procesos se ponga en duda; y así mismo, la legitimidad puede recuperarse si la teoría que buscaba remplazar a la original se comprueba errónea.

Utilidad

La utilidad indica que los resultados de una investigación son aplicables en la industria, muestran la evolución de la ciencia en la tecnología, de la teoría a la práctica.

La utilidad no requiere de la legitimidad, sin embargo, se ve beneficiada por ella, ya que ante mayor aceptación por parte de la comunidad científica y académica, es más probable que los resultados de una investigación se pongan en práctica.

Popularidad

La popularidad indica el interés académico en un tema o investigación, la cantidad de veces que se cita, las ocasiones que se ha intentado replicar. La popularidad depende indirectamente de la legitimidad, sin embargo, la legitimidad no depende de la popularidad; de hecho, la popularidad se puede ganar de forma independiente a la legitimidad y a la utilidad, por ejemplo, cuando nace un interés o curiosidad por un tema nuevo, el cual aún no ha sido investigado, y por lo tanto, ni legitimado ni utilizado en la práctica. Así mismo, la popularidad puede perderse

debido a cambios culturales y generacionales, a pesar de que las investigaciones mantengan legitimidad y utilidad; un ejemplo extremo son las teorías que después de años sin objeción se convierten funcionalmente en leyes, si la academia no considera que puede aportar más, no importa su grado de legitimidad o su historial de uso en la práctica, perderán popularidad.

DISCUSIÓN

Al ser la validez un concepto de tanta importancia para la ciencia, prueba de su mérito ante la sociedad, independientemente de argumentos, la academia difícilmente decidirá desecharla del todo. Por otra parte, las perspectivas epistemológicas, la precisión de sistemas matemáticos, la popularidad académica, son todos temas que invitan a la argumentación y contra-argumentación, y esta disertación no tiene ni busca tener la última palabra, al contrario, se presenta como invitación para continuar analizando la naturaleza y características de estos temas.

Sin embargo, esta disertación sugiere un beneficio para los académicos quienes evalúan la validez de las investigaciones que utilizan como referencia para las investigaciones propias, así como su tiempo de vida útil. Y quizá aún más importante, para el historial de publicaciones académicas de cada investigador, de la cual podría estar en duda su estado de obsolescencia.

Esta disertación quizá tenga un valor limitado para las humanidades y ciencias sociales, las cuales ya han abandonado el concepto de validez, y posiblemente será mirada con recelo por parte de las ciencias duras, las cuáles basan la mayoría de su autoridad en los métodos cuantitativos aquí cuestionados; sin embargo, se sugiere de especial interés para las ciencias económicas y administrativas, las cuales encontrándose en medio de las ciencias duras y blandas, en medio de lo cuantitativo y cualitativo, suelen balancear los conceptos de generalización y especificación, objetividad y subjetividad, por lo que las conclusiones aquí presentadas no les resultan tan disruptivas.

No obstante, lo principal es invitar a los investigadores cuantitativos a cuestionarse acerca del concepto de validez, concepto que en muchas ocasiones se vuelve una obsesión para el académico, y que sin importar el método, sin importar la precisión de los instrumentos y aplicación, sin importar la amplitud de las variables analizadas, nunca se alcanzará a la perfección.

CONCLUSIONES

El concepto de validez es un cimiento de la ciencia, es lo que la diferencia de mitos, costumbres y rumores, sin embargo, el constante cambio de lo que se considera válido o inválido, la evolución de las herramientas que lo definen, muestran de una forma práctica la debilidad de este cimiento.

Por otra parte, desde una perspectiva epistemológica, quizá resulte imposible el asegurar de forma determinante y definitiva la validez total de cualquier investigación, por lo menos hasta que las herramientas necesarias para comprobarla, los sistemas matemáticos, se desarrollen al punto de la perfección.

Además, la academia suele priorizar el interés y la popularidad ante la validez de las investigaciones, y las llamadas ciencias "blandas", en algunos casos han decidido darle la espalda, acogiendo la variabilidad y subjetividad de sus resultados.

Sin embargo, a pesar de que la validez no es un concepto perfecto, la legitimidad que le confiere a las investigaciones motiva a que éstas se lleven a la práctica, lo que demuestra su utilidad al probarlas ante la realidad, y al fomentar su popularidad, llama la atención de nuevos académicos a que se unan a las investigaciones, mejorando el alcance y la profundidad del conocimiento, y en algunas ocasiones, actualizándolo. No obstante, resultaría conveniente detener el uso de un término tan cuestionable y epistemológicamente incorrecto como la validez, para remplazarlo por términos que definan mejor sus características y limitaciones, como la legitimidad, utilidad y popularidad.

¿Y qué debería ocurrir con la investigación que utiliza datos validados mediante la prueba K-S que no resulten validados con Lilliefors? Desde una perspectiva epistemológica, independientemente de la prueba K-S o Lilliefors, la investigación carece del concepto de validez absoluta y dogmática. Sin embargo, hablando en términos de legitimidad académica, si se le había conferido legitimidad bajo los estándares aceptables en el momento de ser aceptada, debería mantener dicha legitimidad hasta que una nueva investigación la contradiga, o hasta que por elementos externos (políticos, ecológicos, culturales, generacionales, etc...) cambie la realidad que dicha investigación buscaba reflejar. En ese caso, y si el objeto de estudio mantiene popularidad en la academia, resultaría provechoso el realizar una actualización, tanto de los datos como del método, para entonces revalidar la investigación, evitando la obsolescencia.

REFERENCIAS

1. Adetunji, O., Bischoff, J., & Willy, C. J. (2018). Managing system obsolescence via multicriteria decision making. *Systems Engineering*, 21*(4)*, 307–321. doi:10.1002/sys.21436
2. Arao, L. H., Veloso, M. J. & da Silveira, V. L. (2015). The Half-Life and Obsolescence of the Literature Science Area: a contribution to the understanding the chronology of citations in academic activity. *Qualitative and Quantitative Methods in Libraries, 4, 603-610. Recuperado de qqml.net*
3. Avis, M. (1995). Valid arguments? A consideration of the concept of validity in establishing the credibility of research findings. *Journal of Advanced Nursing*, 22*(6)*, 1203–1209. doi:10.1111/j.1365-2648.1995.tb03123.x
4. Burton, R. E., & Kebler, R. W. (1960). The "half-life" of some scientific and technical literatures. *American Documentation*, 11*(1)*, 18–22. doi:10.1002/asi.5090110105
5. Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The Concept of External Validity. *Journal of Consumer Research, 9(3), 240-244. Recuperado de jstor.org*
6. Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54*(4)*, 297–312. doi:10.1037/h0040950
7. Cunningham, S. J., & Bocock, D. (1995). Obsolescence of computing literature. *Scientometrics*, 34*(2)*, 255–262. doi:10.1007/bf02020423
8. Hayashi, P., Abib, G., & Hoppen, N. (2019). Validity in Qualitative Research: A Processual Approach. *The Qualitative Report*, 24*(1)*, 98-112. Recuperado de: nova.edu
9. Hessen, J. (1926). *Erkenntnistheorie: Leitfäden der Philosophie, 2. Bd*. Berlin: F. Dümmler
10. IBM. (2018). K-S test of normality in NPAR TESTS and NPTESTS does not use Lilliefors correction prior to SPSS Statistics 21.0.0.1, *IBM Support*. Recuperado de ibm.com
11. Kirchgässler, K. U. (1991). Validity ? the quest for reality in quantitative and qualitative research. *Quality & Quantity*, 25*(3)*, 285–295. doi:10.1007/bf00167533

12. Koch, T. (1994/2006). Establishing rigour in qualitative research: the decision trail. *Journal of Advanced Nursing*, 53*(1)*, 91–100. doi:10.1111/j.1365-2648.2006.03681.x

13. Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, 62*(318)*, 399. doi:10.2307/2283970

14. Line, M.B. (1993). Changes in the Use of Literature with Time--Obsolescence Revisited. *Library Trends*, 41, 665-683. Recuperado de semanticscholar.org

15. Meijer, P. C., Verloop, N., & Beijaard, D. (2002). Multi-Method Triangulation in a Qualitative Study on Teachers' Practical Knowledge: An Attempt to Increase Internal Validity. *Quality and Quantity*, 36*(2)*, 145–167. doi:10.1023/a:1014984232147

16. Oberhofer, C. M. A. (1993). Information use value: A test on the perception of utility and validity. *Information Processing & Management*, 29*(5)*, 587–600. doi:10.1016/0306-4573(93)90081-n

17. Onwuegbuzie, A. J., Johnson, R. B. (2006). The Validity Issue in Mixed Research. *Research in the Schools. 13(1), 48-63. Recuperado de researchgate.com*

18. Russell, E. W. (2010). The "Obsolescence" of Assessment Procedures. *Applied Neuropsychology*, 17*(1)*, 60–67. doi:10.1080/09084280903297917

19. Smirnov, N. V. (1939). Estimate of deviation between empirical distribution functions in two independent samples. *Bull Moscow University*, 2*(2), 3-16.*

20. *Song, Y., Ma, F., & Yang, S. (2014). Comparative study on the obsolescence of humanities and social sciences in China: under the new situation of web. Scientometrics, 102(1), 365–388. doi:10.1007/s11192-014-1410-8*

21. Van Aken, J. E. (2005). Management Research as a Design Science: Articulating the Research Products of Mode 2 Knowledge Production in Management. *British Journal of Management,* 16(1), 19–36. doi:10.1111/j.1467-8551.2005.00437.x

22. Van Horn, M. L., Smith, J., Fagan, A. A., Jaki, T., Feaster, D. J., Masyn, K., Hawkins, J. D., & Howe, G. (2012). Not Quite Normal: Consequences of Violating the Assumption of Normality in Regression Mixture Models. Structural Equation Modeling. *A Multidisciplinary Journal,* 19(2), 227–249. doi:10.1080/10705511.2012.659622

23. Weber, M. (1922/1968). *Economy and Society.* Berkeley, California: University of California Press

24. Zimmerman, D. W. (1998). Invalidation of Parametric and Nonparametric Statistical Tests by Concurrent Violation of Two Assumptions. *The Journal of Experimental Education,* 67(1), 55-68. doi: 10.1080/00220979809598344

CONSIDERACIONES EN LAS METODOLOGÍAS CUANTITATIVAS PARA CIENCIAS ECONÓMICO-ADMINISTRATIVAS CON USO DE REGRESIÓN LINEAL MÚLTIPLE.

Gonzalo R. Ceballos
Dr. Victor Manuel Larios Rosillo

Palabras Clave: Regresión lineal múltiple, supuestos estadísticos, técnicas multivariantes.

INTRODUCCIÓN

A principios de siglo pasado, se consolidó la visión objetivista y positivista de la ciencia a través de las teorías y nociones de certidumbre y probabilidad junto con la validación de hipótesis, procesos de estandarización y los enfoques experimentales (Marquez, E. 2013). Tal visión epidemiológica fue asumida por las ciencias sociales y es expresada a través de la investigación cuantitativa en sus diversos modelos de expresión científica, lo cual ha contribuido en la fidelidad y normatividad de las metodologías que dan sustento a tal rama de la ciencia misma, representado principalmente por los métodos de investigación cuantitativos.

Para ello han surgido (y adaptado) diversa técnicas de fundamentos positivistas basados en modelos matemáticos que nos permiten comprender la realidad y en ocasiones predecirla siempre y cuando se cumplan con supuestos estadísticos específicos y exista un respaldo teórico que fundamente los procesos de la investigación. Poole, M., & O'Farrell, P. 1971; Woodward J. 1998)

El objetivo (en la mayoría de los casos) de las investigaciones cuantitativas es el de analizar cómo la manipulación o cambio de una o más variables (o factores) afecta la varianza de uno o más indicadores de rendimiento de otra(s) variable(s), en donde las

variables manipuladas son llamadas "variables independientes" y las afectadas "variables dependientes", para tales casos, cuando la intención es la de explorar y predecir relaciones, existen los análisis de regresión (Nelson, L. et al.,1979). La relevancia de el presente trabajo consta de la conglomeración de un panorama general de los aspectos más relevantes en temáticas de regresión lineal múltiple y los cuidados (y consideraciones) a considerar para generar un desempeño favorable y de buen rendimiento en las investigaciones cuantitativas en ciencias sociales y económicas.

DESARROLLO

Dependiendo la nvestigación se darán situaciones en las que el estudio precisa de solo 2 variables para formar y analizar su proceso metodológico, sin embargo, existen situaciones más complejas que suelen constar de la relación de 3 o más variables. A las técnicas utilizadas para llevar a cabo este tipo de análisis se les conoce como técnicas multivariantes. En tales casos, cuando se quiere definir y determinar las relaciones de dichas variables con posibles eventos en el futuro se aprovechan las bondades de las técnicas de regresión múltiple. (Cardona, F., Gonzalez, L., Rivera, M., Cárdenas, E. 2013).

Contexto histórico y práctico

A finales del siglo XIX, surge el concepto de regresión múltiple fundamentado con los trabajos de Francis Galton (1877) y fueron rápidamente continuados por Edgeworth (1893), Pearson (1896), Filon (1898), Yule, (1911) y posteriormente Fisher (1922) quien sintetizó el trabajo de los autores previos y creó un modelo moderno de regresión múltiple al cual posteriormente apoyó Barlett (1933) en su texto "On the theory of statistical regression" (Aldrich, J. 2005) en el cual no solo se fundamentan herramientas de análisis de regresión múltiple, sino también pruebas estadísticas de fiabilidad y validez de modelos a través de fundamentos matemáticos cuyo propósito original fue el de dar robustez y consolidar las técnicas de regresión lineal.

Estas herramientas fueron cuestionadas a mediados del siglo pasado por la comunidad científica enfocada en desarrollo de investigaciones de carácter experimental a través de ANOVA/ANCOVA (técnicas que surgieron a la par de las regresiones), sin embargo con el paso del tiempo y la adaptación a sistemas computacionales de gran poder estadístico, esta crítica cesó debido a los notables alcances de los sistemas de regresión (Cohen, J. & Cohen, P. 2003).

Los análisis de regresión han sido unas de las principales técnicas multivariantes utilizadas por una gran variedad de disciplinas tales como la ecología, ciencias sociales y económicas (Wagner, H., 2013; Aronow, P., & Samii, C. 2016) y esto es "fácil de percibir" en el ambiente científico dada su presencia en una basta cantidad de artículos e investigaciones científicas presentes en revistas mundialmente reconocidas y planes de estudio de posgrados.

Por ejemplo, en temas socio-económicos, una de las aplicaciones estadísticas multivariantes de regresión alabadas es la de England, Farkas, Kilbourne & Dou (1988) citada en Cohen, J. & Cohen, P. (2003) donde se pudo predecir que el tamaño de la relación positiva entre el número de años de experiencia de los trabajadores y su salario dependería de la ocupación de trabajadoras femeninas en la organización, un hallazgo llamativo para la época. Así mismo ha sido usada también en ámbitos legales en donde en Estados Unidos es considerada como una prueba efectiva de litigio en el ámbito legal desde 1964 (Law and Contemporary Problems, 1983).

Bondades de la técnica.

Cohen J. (1968) señala que las técnicas relacionadas a la regresión múltiple constan de un gran poder estadístico, flexibilidad y fidelidad, así mismo la definen como una herramienta con un poder predictivo óptimo y utilidad de determinación causal entre una buena gama de variables, por ello su adaptación a las ciencias sociales pudiera parecer incluso una sucesión de adaptación natural o esperada.

Se puede definir a la regresión lineal múltiple como un proceso para evaluar las relaciones entre varias variables independientes con respecto a una variable dependiente. Dichas técnicas tienen similaridades con las ANOVAS Y ANCOVAS, sin embargo el uso de la regresión múltiple ha tenido beneficios populares debido a su aceptado poder predictivo y su versatilidad al tratar con una gran variedad de variables dependientes, pudiendo ser estas categóricas o continuas. (Nelson, L. et al.1979),

Incluso es mencionado que los datos analizados a través de regresión lineal múltiple son compatibles y analizables a través de técnicas como ANOVAS, pero esto no es posible en el modo inverso debido a no-idependencia de los factores de la información que se explora en los modelos de regresión (Cohen, J. & Cohen, P. 2003), lo cual expresa ciertas limitantes del Analisis de Varianza.

Las técnicas de regresión permiten afirmar ciertos supuestos sobre conexiones causales entre variables para usar información sobre varianzas y covarianzas en la prueba de otras afirmaciones (Woodward J. 1998). Son estos atributos los que inclinan a tales técnicas a inferir predicciones, siempre y cuando se cumplan los supuestos matemáticos particulares requeridos para su correcta aplicación.

Desglose teórico y requerimientos.

Las regresiones múltiples pueden ser usadas en cualquier caso en el que una variable cuantitativa (y) sea objeto de estudio en su relación con varios factores de interés igualmente cuantitativos que funcionen como variables independientes, de tal manera que se pueden representar el tipo de relaciones entre ellas y la complejidad misma con "fidelidad" (Cohen, J. & Cohen, P. 2003) a través de la siguiente representación en términos conceptuales:

$$Y = a + bU + cV + dW + eX + \cdots$$

En la cual Y representa a la variable dependiente, el resto de las mayúsculas (U,V,W,X…) representan las variables independientes que varían en cantidades representadas en minúsculas (b, c, d, e…) con la constante "a". Tal representación conceptual nos permite expresar que la variable Y (dependiente) se compone de la suma y varianza de las variables independientes para fines explicativos. De igual manera, profundizando en la naturaleza matemática de la técnica, se puede tener un mejor alcance explicativo a través de su expresión matemática básica:

$$Y = a + \sum_{i=1}^{k} b_i X_i + u$$

Donde Y representa la variable dependiente conformada por los valores equivalentes de la suma de "k" número de variables independientes "X" y su relación con los coeficientes de regresión "a" y "b" representando parámetros del modelo para una población específica; sumando "u" que representa un término de perturbación escolástica como efecto de posibles variables omitidas (Poole, M., & O'Farrell, P. 1971; Johnston, J. 1963; Cohen, J. & Cohen, P. 2003).

Como elemento a considerar, Thomas, D et. Al. (2007) mencionan que para que los análisis de regresión tengan un sustento estadístico de confianza aceptable, son requeridas 250 o más muestras basados en la teoría de Pratt, ya que con muestras menores puede haber alteraciones en los resultados y su valor estadístico. Así mismo, para autenticar el uso de una herramienta de regresión múltiple se requiere que la información poseída cuente con los siguientes supuestos técnicos (Poole, M., & O'Farrell, P. 1971):

- Cada valor de X y Y se observa sin error de medición.
- La relación entre Y y cada variable X es lineal en los parámetros seleccionados de la función.

- Cada distribución condicional de "u" tienen una media de 0.
- Se asume homosedasticidad los valores en relación con "u".
- La varianza en la distribución condicional de "u" es constante.
- Los valores de "u" son independientes entre sí con una covariaza de 0.
- La variables independientes X son independiente a su vez entre sí.
- La distribución de las variables es una distribución normal.

DISCUSIÓN

Resulta importante mencionar que respecto a la selección de variables independientes con relación a la variable dependiente, es importante considerar que en el análisis de regresión (tanto como en otras técnicas de correlación) debemos asumir que solo hay unos pocos factores de "confusión" a revelar, de lo contrario no podremos detectar la estructura causal subyacente (Bartelborth, T. 2011) esto haciendo alusión al concepto de parsimonía. Por ello se debe partir de la teoría existente, ya que de usar estas técnicas como meramente un análisis exploratorio, pondrá en riesgo la veracidad de la investigación.

Por su parte, Woodward J. (1998) menciona que en términos generales, en lo que respecta a temas de regresión lineal (en general) existen ciertas máximas o afirmaciones aceptadas, tales como que las regresiones lineales son técnicas para la creación de inferencias causales en circunstancias en las que el investigador carece de conocimiento suficientes sobre "leyes generales" o teorías sistematizadas sobre un fenómeno; que las regresiones lineales (al igual que otros métodos de modelado técnico) no son viables para generar afirmaciones causales sobre patrones de generalidad estadística; las afirmaciones causales de las re-

gresiones funcionan como afirmaciones a nivel de población, y no necesariamente se pueden extender a realizar afirmaciones sobre individuos particulares.

Vale la pena también mencionar un aspecto fundamental sobre el uso de ésta (y otras) herramientas para generar conocimiento, y es que a pesar el ser humano siempre ha estado en la búsqueda de conocimientos en la realidad; no obstante, en las ciencias sociales dicha realidad cambia constantemente (Del Canto, E., Silva, A. 2013), por ello Achen, C. (1982) mencionaba claramente que si bien pueden resultar exitosas las conclusiones de una regresión, no se pueden generalizar como leyes o normas naturales, pues estas pueden cambiar con el tiempo. Por ello, el contexto teórico de la investigación deberá tener un valor notable para validar los resultados. Así mismo, en el caso de el estudio aplicado a ciencias económico-administrativas, deberá tomar en cuenta el factor constante de cambio del entorno y el objeto de estudio en sí.

CONCLUSIONES

Haciendo una recapitulación de los puntos y áreas mencionadas en el documento podemos llegar a las siguientes conclusiones: 1. Las herramientas de regresión lineal múltiple tienen un fuerte respaldo teórico junto con una flexibilidad y adaptación que hace posible aplicarlas en contextos variados de ciencias positivistas, naturales, sociales y económicas. 2. Si bien tienen un fundamento estadístico sólido y cuentan con una metodología propia para validar el funcionamiento de estas bajo supuestos establecidos, se debe cuidar que la teoría y el contexto que respalde a los procesos estadísticos como recomendación fundamental, por ello, a pesar de la fiabilidad matemática de los resultados, no deberían utilizarse estas herramientas como único medio para validad leyes o "teorías" fundamentales como menciona Woodward J.(1998). 3. Sus desventajas respecto a la posible omisión de variables relacionadas deja puerta a enriquecer las investigaciones que usen dichas herramientas con complementación con otras técnicas y fundamentos que permitan complementar y respaldar los resultados así como validar la veracidad de los mismos.

REFERENCIAS

1. Achen, C. (1982). Interpreting and Using Regression. Beverly Hills: Sage Publications.
2. Aldrich, J. (2005). Fisher and Regression. *Statistical Science,* 20(4), 401-417. Retrieved from http://www.jstor.org/stable/20061201
3. Aronow, P., & Samii, C. (2016). Does Regression Produce Representative Estimates of Causal Effects? *American Journal of Political Science*, 60(1), 250-267. Retrieved from http://www.jstor.org/stable/24583062
4. Bartelborth, T. (2011). Propensities and Transcendental Assumptions. *Erkenntnis* (1975-), 74(3), 363-381. Retrieved from http://www.jstor.org/stable/41476694
5. Cardona, F., Gonzalez, L., Rivera, M., Cárdenas, E. (2013). Inferencia estadística: Módulo de regresión lineal simple. ISSN: 0124-8219.
6. Cohen, J. (1968) Multiple regression as a general data-analytic system. Psychological Bulletin.
7. Cohen, J. & Cohen, P. (1975) Applied multiple regression I correlation analysis for the behavioral sciences. Hillsdale, N.J.: Lawrence Eribaum Assoc. Pub.
8. Dana, J., & Dawes, R. (2004). The Superiority of Simple Alternatives to Regression for Social Science Predictions. *Journal of Educational and Behavioral Statistics,* 29(3), 317-331. Retrieved from http://www.jstor.org/stable/3701356
9. Del Canto, Ero, & Silva Silva, Alicia (2013). Metodología cuantitativa: Abordaje desde la complementariedad en ciencias sociales. Revista de Ciencias Sociales (Cr), III(141), undefined-undefined. [fecha de Consulta 12 de Noviembre de 2019]. ISSN: 0482-5276. Disponible en: https://www.redalyc.org/articulo.oa?id=153/15329875002
10. HINDMAN, M. (2015). Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. *The Annals of the American Academy of Political and Social Science,* 659, 48-62. Retrieved from http://www.jstor.org/stable/24541848
11. JOHNSTON, J. (1963). Econometric methods: An introduction to linear statistical models. 5-6; F. A. GRAYBILL, vol. I

12. Karakostas, K. (2004). Interpreting Regression Diagnostics. Journal of Educational and Behavioral Statistics, 29(3), 369-373. Retrieved from http://www.jstor.org/stable/3701359

13. Márquez, E. (2013). La perspectiva epistemológica objetivista y la hegemonía de la investigación cuantitativa en las ciencias sociales. Revista de Investigación, 37(78), undefined-undefined. [fecha de Consulta 12 de Noviembre de 2019]. ISSN: 0798-0329. Disponible en: https://www.redalyc.org/articulo.oa?id=3761/376140393001

14. Nelson, L., Nelson, L., & Zaichkowsky, L. (1979). A Case for Using Multiple Regression Instead of ANOVA in Educational Research. The Journal of Experimental Education, 47(4), 324-330. Retrieved from http://www.jstor.org/stable/20151298

15. Poole, M., & O'Farrell, P. (1971). The Assumptions of the Linear Regression Model. Transactions of the Institute of British Geographers, (52), 145-158. doi:10.2307/621706

16. Thomas, D., Zhu, P., & Decady, Y. (2007). Point Estimates and Confidence Intervals for Variable Importance in Multiple Linear Regression. Journal of Educational and Behavioral Statistics, 32(1), 61-91. Retrieved from http://www.jstor.org/stable/20172069

17. Title VII, Multiple Linear Regression Models, and the Courts: An Analysis. (1983). Law and Contemporary Problems, 46(4), 283-295. doi:10.2307/1191603

18. Wagner, H. (2013). Rethinking the linear regression model for spatial ecological data. Ecology, 94(11), 2381-2391. Retrieved from http://www.jstor.org/stable/23597200

19. Woodward, J. (1988). Understanding Regression. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1988, 255-269. Retrieved from http://www.jstor.org/stable/192993

LÓGICA DIFUSA, REGRESIÓN MÚLTIPLE, RED NEURAL ARTIFICIAL PARA SU USO EN LAS CIENCIAS DE LA ADMINISTRACIÓN

Ricardo de Jesús Nuño Velasco
Dr. Juan Mejía Trejo

Palabras Clave: Lógica Difusa, Regresión Múltiple, Red Neuronal Artificial.

INTRODUCCIÓN

La Lógica Difusa (LD) es un formalismo matemático que pretende emular la habilidad que tienen algunas personas para tomar decisiones correctas a partir de datos vagos o imprecisos y que están expresados lingüísticamente. Permite tratar información imprecisa, como estatura media o temperatura baja, en términos de conjuntos difusos que se combinan en reglas para definir acciones, por ejemplo, "si la temperatura es alta entonces enfriar más" (Klir, et al. 1997).

Sin embargo, según Zimmermann (2001), se recomienda el uso de reglas poco complejas; la selección pragmática de operadores, que en combinación con la defusificación, sólo permite buenos resultados con reglas simples.

Una aplicación de la LD en la electrónica es la de medir los niveles de Radio Frecuencia (RF) a la que estamos expuestos con el uso de teléfonos móviles, ya que los niveles recomendados por la International Commission on Non-ionizing Radiation Protection (ICNIRP), están regulados por los fabricantes. Sin embargo según Bit-Babik, et al. (2003), la cantidad de energía de radiofrecuencia RF a la que una persona está expuesta depende de muchos factores.

La Lógica Difusa Compensatoria (LDC), según Cejas et al. (2012) ha sido utilizada en el sector empresarial en diversos países como Alemania, España, Argentina, Brasil y Cuba. Se destacan aplicaciones recientes en este ámbito en la selección de

proveedores, soluciones clásicas de juegos cooperativos y para determinar la confiabilidad de una empresa en términos económicos. Por lo anterior, se destaca la importancia del presente ensayo en realizar una comparación de usos de las tres técnicas.

DESARROLLO

A continuación, se tratarán los conceptos básicos de las tres técnicas a fin de determinar un cuadro comparativo.

Lógica Difusa

Según Zadeh (1965), la LD emplea una terminología particular: "*Fuzzy*" (Difuso o Borroso); "*Crisp*" (Nítido); "Fusificación" (convertir un conjunto nítido en borroso); y "Defusificación" (convertir un conjunto borroso en un valor nítido).

Los conjuntos difusos pueden ser considerados como una generalización de los conjuntos clásicos: la teoría clásica de conjuntos solo contempla la pertenencia o no pertenencia de un elemento a un conjunto, sin embargo la teoría de conjuntos difusos contempla la pertenencia parcial de un elemento a un conjunto, es decir, cada elemento presenta un grado de pertenencia a un conjunto difuso que puede tomar cualquier valor entre 0 y 1 Mendoza (2017), ver Tabla 1.

Tabla 1. Valores de Verdad.

Valor de Verdad	Categoría
0	falso
0,1	casi falso
0,2	bastante falso
0,3	algo falso
0,4	más falso que verdadero
0,5	tan verdadero como falso
0,6	más verdadero que falso
0,7	algo verdadero
0,8	bastante verdadero
0,9	casi verdadero
1	verdadero

Mendoza (2017)

Para Kecman (2001), el cuerpo teórico constituye una rama de la Inteligencia Computacional que se funda en la incertidumbre, lo cual permite manejar información vaga o de difícil especificación si se quiere utilizar objetivamente esta información con un fin específico.

Según Dubois et al. (1980), las distintas formas de definir las operaciones y sus propiedades determinan diferentes lógicas multivalentes que son parte del paradigma de la LD.

Regresión Múltiple

La regresión múltiple es una técnica estadística que permite determinar la correlación que existe entre variables independientes y dos o más variables dependientes. La regresión múltiple se puede utilizar para analizar datos ordinales y categóricos (Kalyan & Choudhury 2008).

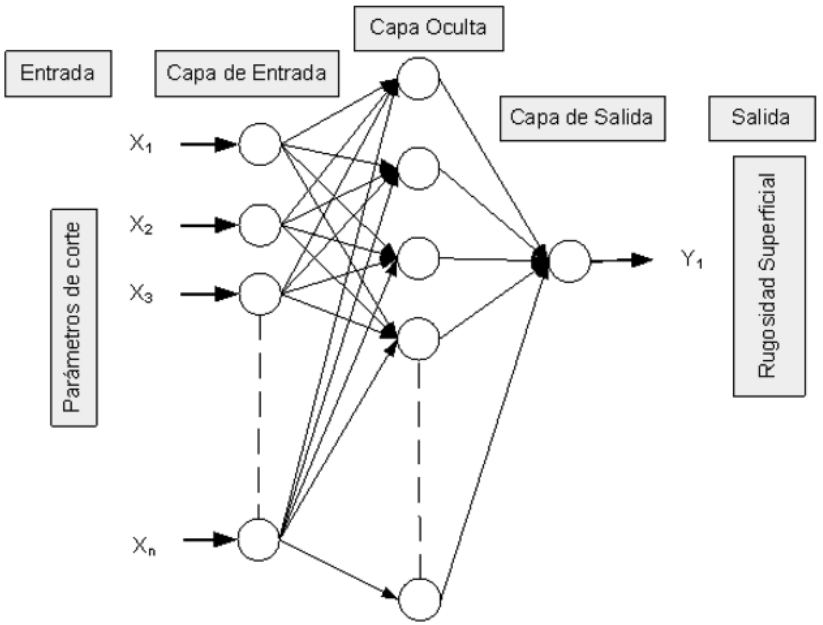
Por lo general, según Montgomery (2002) se realiza primeramente un análisis de varianza (Anova) para determinar los

factores importantes involucrados y luego con el uso de la regresión se obtiene un modelo cuantitativo que relaciona los factores más importantes con la respuesta.

Red Neuronal Artificial

Para Morales et al. (2018) en una red neuronal artificial, la determinación del número óptimo de neuronas de la capa oculta se realiza mediante un proceso de ensayo y error en el que se prueban diferentes variantes. En todo caso, el objetivo es dotar a la red de un número adecuado de neuronas en la capa oculta para garantizar la capacidad de aprendizaje de las características de las posibles relaciones existentes entre los datos de la muestra ver Figura 1.

Figura 1. Estructura de la red perceptrónica multicapa.



Morales et al. (2018)

DISCUSIÓN

La propuesta Metodológica del Análisis de Datos Borrosos, involucra 4 grandes procesos: 1- La codificación y estandarización de los datos medidos; 2- La definición de las variables lingüísticas difusas y su empleo en la transformación de los datos en números borrosos representados en tablas de contingencia; 3- La obtención de un valor borroso colectivo y un valor nítido final representativo de la calificación de las muestras evaluadas; 4- La aplicación de métodos multivariados para el análisis, visualización y obtención de conocimiento sobre los datos borrosos Césaria et al. (2017).

Según Trillas & Gutiérrez (1992), para que se cumpla la equivalencia entre lógica clásica y lógica difusa se debe traducir por una t-norma, ya que preserva la relación causa-efecto y el sentido físico.

Según Guadarrama (2000), se pueden distinguir tres clases de sistemas basados en lógica difusa de acuerdo con la forma de las reglas y el tipo de entradas y salidas:

a. **Sistemas difusos tipo Mamdani**: Están compuestos por una base de conocimiento, un motor de inferencias, un bloque de fusificación y un bloque de desfusificación.
b. **Sistemas difusos puros**: Estos sistemas tienen como entrada y como salida conjuntos borrosos. Al no realizar ninguna transformación sobre las entradas o sobre las salidas, tienen solo dos componentes principales: una base de conocimiento y un motor de inferencias.
c. **Sistemas difusos tipo Takagi-Sugeno-Kang**: En lugar de trabajar con reglas lingüísticas, Takagi, Sugeno y Kang propusieron un nuevo modelo basado en reglas donde el antecedente estaba compuesto de variables lingüísticas y el consecuente se representaba como una función de las variables de entrada Schnitman y Yoneyama (2001).

Según Ríos (2005), aunque existen distintos métodos de desfusificación, uno de los más usados en aplicaciones en gestión es el Centro de Sumas. Este considera la contribución individual del área de cada conjunto de salida formado al aplicar cada una de las reglas:

$$x_{salida} = \frac{\sum_{i=1}^{l} x_i \sum_{k=1}^{n} \mu^{(k)}(x_i)}{\sum_{i=1}^{l} \sum_{k=1}^{n} \mu^{(k)}(x_i)}$$

Cuando el método de LD se compara contra otro método estadístico como lo es la regresión lineal, para resolver un problema de selección de personal, según el estudio realizado por Díaz et al. (2014), los resultados muestran que el modelo optado utilizando variables difusas triangulares y con solapamiento de 25% del área es ampliamente superado por el modelo de regresión múltiple, sería recomendable probar otros modelos de lógica difusa.

Por lo tanto, para Cejas (2011), en los procesos que requieren toma de decisiones, el intercambio con los expertos lleva a obtener formulaciones complejas y sutiles que requieren de predicados compuestos. Los valores de verdad obtenidos sobre estos predicados compuestos deben poseer sensibilidad a los cambios de los valores de verdad de los predicados básicos. Esta necesidad se satisface con el uso de la LDC, que renuncia al cumplimiento de las propiedades clásicas de la conjunción y la disyunción, contraponiendo a éstas la idea de que el aumento o disminución del valor de verdad de la conjunción o la disyunción provocadas por el cambio del valor de verdad de uno de sus componentes, puede ser "compensado" con la correspondiente disminución o aumento de la otra.

La Lógica Difusa Compensatoria (LDC) según Espín et al. (2011), se trata de un nuevo sistema multivalente que rompe con la axiomática tradicional de este tipo de sistemas para lograr un comportamiento semánticamente mejorado respecto a los sistemas clásicos de LD.

Delgado (2005), asegura que la LDC ha sido utilizada en problemas diversos de toma de decisiones destacándose su aplicación en pequeñas y medianas empresas (PYMES) argentinas y brasileras de diferentes esferas de producción y servicios. En este campo sobresalen Espín y Vanti (2005), al aplicar un modelo de LDC para el análisis de la planeación estratégica de una empresa de Comercio Exterior de la Región de Rio Grande del Sur en Brasil.

Por otra parte, las redes neuronales artificiales (RNA) son ampliamente usadas en muchas aplicaciones de la industria. Estas son muy populares en la modelación de sistemas debido a su alta eficiencia en la adaptación y en el aprendizaje mediante el reconocimiento de patrones (Mia y Dhar 2016).

En el estudio realizado por Morales et al. (2018), la red instalada en esta investigación es una red perceptrónica multicapa la cual corresponde en equivalencia a la regresión no lineal múltiple. La red perceptrónica multicapa está compuesta por la asociación de neuronas artificiales organizadas dentro de la red formando niveles o capas, el entrenamiento fue desarrollado a través del algoritmo Levenberg Marquardt. Los mejores resultados fueron obtenidos con la estructura 3-5-1, tres neuronas en la capa de entrada, 5 neuronas en la capa oculta y una en la capa de salida. El software de redes neuronales fue codificado utilizando el Neural Networks Toolbox de Matlab (Montaño, 2002).

Sin embargo, para Morales et al. (2018), en su estudio de comparación entre una red neural artificial (RNA) y una regresión múltiple, los menores errores medios absolutos fueron obtenidos con los modelos implementados con redes neuronales artificiales.

Para ver cómo se pueden comparar los distintos métodos antes mencionados ver Tabla 2.

Tabla 2. Comparación de los métodos LD, LDC, Regresión Múltiple y Red Neural Artificial.

MÉTODO	DEBILIDAD	FORTALEZA	AUTOR
Lógica Difusa	Reglas poco complejas	Puede trabajar con datos borrosos	Zadeh (1965)
Regresión Múltiple	Requiere Linealidad y normalidad	Tiene una confiabilidad comprobada	Alonso (2007)
Lógica Difusa Compensatoria	No es tan precisa como otros métodos estadísticos	Puede trabajar con conjuntos complejos	Cejas (2011)
Red Neural Artificial	No son estables, requieren un entrenamiento preciso.	Alta eficiencia y adaptación por aprendizaje.	Mia y Dhar (2016)

"Elaboración Propia"

CONCLUSIONES

Una gran desventaja de la RNA es que requieren de software de alto nivel "Networks Toolbox de Matlab" como el que utilizo Morales et al. (2018) en su estudio, otras alternativas son el SPSS (Statistics o Modeler) de IBM y el Orange software libre que también cuenta con el módulo de Red Neuronal Artificial.

No cabe duda que las nuevas técnicas de predicción pueden llegar a ser más precisas que las tradicionales, pero requieren de una serie de pruebas y una configuración más rigurosa que los métodos tradicionales como la regresión múltiple, además, el grado de precisión de los métodos tradicionales no ha sido ampliamente superados por la RNA y la LD a tal grado que llegan a ser mejores en algunas circunstancias.

Se puede concluir que la mejor manera de aplicar estas nuevas técnicas de análisis estadístico, es mediante su uso combinado con técnicas tradicionales como la Regresión Múltiple, ya que estas permiten tener una mejor interpretación y validación de los resultado obtenidos en nuestros estudios futuros.

REFERENCIAS

1. Alonso, J. (2007). Tutorial para la estimación de un modelo de regresión múltiple e inferencia con EasyReg. St. Louis: Federal Reserve Bank of St Louis.
2. Bit-Babik, G.; Chou, C.K.; Faraone, A.; Gessner, A.; Kanda, M.; Balzano, P. (2003). "Estimación de la SAR en la cabeza humana y cuerpo debido a la exposición a la radiación radiofrecuencia de teléfonos móviles con accesorios de manos libres". Radiat Res Vol. 159 N.° 4, pp. 550-557.
3. Cejas, J., Espín, R. y Alfonso, D. (2012). "Aplicación de la lógica difusa Compensatoria en el sector empresarial". Bilbao, España: DYNA, Vol. 87, N.° 3.
4. Cejas-Montero, J. (2011). "LA LÓGICA DIFUSA COMPENSATORIA". Ingeniería Industrial, Vol.32, N.° 2, ISSN: 02585960.
5. Césaria, M. I., Gámbarco, A., y Césaric, R. (2017). Metodología de análisis de datos imprecisos con lógica difusa. REVISTA ARGENTINA DE INGENIERÍA, Vol. 9.
6. Delgado, T. (2005). "Capacity- building: spatial data infrastructure readiness index". Proceedings of 8th UN Regional Cartographic Conference for the Americas New York.
7. Díaz, C.; Aguilera, A.; Guillén, N. (2014). "Lógica Difusa vs. Modelo de Regresión Múltiple para la selección de personal".
8. Dubois, D.; Prade, H. (1980). "Fuzzy Sets and Systems: Theory and Applications". Academic Press, New York, ISBN: 9780122227509.
9. Espín, R.; Fernández, E. y González, E. (2011). "Un sistema lógico para el razonamiento y la toma de decisiones: Lógica Difusa Compensatoria basada en la media geométrica". Revista de Investigación Operacional., Vol. 32, N.° 3, pp. 230-245.
10. Espín, R.; Vanti, A. (2005). "Administración Lógica: Un estudio de caso en empresa de comercio exterior". Revista Base, Vol. 2, N.° 2, p. 69-77. ISSN 1984-8196.
11. Guadarrama-Cotado, S. (2000). Representación del conocimiento impreciso: Revisión parcial de las teorías de conjuntos borrosos. Doctoral dissertation, Facultad Informática, Universidad Politécnica De Madrid.
12. Kalyan, K y Choudhury, S. (2008). "Investigation of tool wear and cutting force in cryogenic machining using design of experiments," Journal of Materials Processing Technology, vol. 203, N°. 1, pp. 95–101.

13. Kecman, V. (2001). “Learning and Soft Computing-Support Vector Machines, Neural Networks and Fuzzy Logic Models”. Massachusetts: The MIT Press., ISBN: 0262112558.

14. Klir, J., Clair, U. St. y Yuan, B. (1997). “Fuzzy set theory: foundations and applications”. Editorial Prentice Hall.

15. Mendoza, F. (2007). “Redes neuronales y lógica difusa en la predicción del crecimiento de una matrícula estudiantil - docente”. Tesis de grado para optar al título de Licenciatura en Informática mención Ingeniería de Sistemas Informáticos. Universidad Mayor de San Andrés. La Paz, Bolivia.

16. Mia, M y Dhar, N. (2016). “Prediction of surface roughness in hard turning under high pressure coolant using artificial neural network,” Measurement, Vol. 92, Supplement C, pp. 464–474.

17. Montaño, J. (2002). “Redes neuronales artificiales aplicadas al análisis de datos,” Ph.D. dissertation, Universitat de Les Illes Balears. Islas Baleares, España.

18. Montgomery, D. (2002). “Design and Analysis of Experiments (fifth edition)”. John Wiley & Sons, Ltd.

19. Morales, Y.; Zamora, Y.; Vásquez, P.; Porras, M.; Bárzaga, J.; López, R. (2018). “Comparación entre redes neuronales artificiales y regresión múltiple para la predicción de la rugosidad superficial en el torneado en seco”. Ingenius. N°. 19, pp. 79-88.

20. Ríos, J. (2005) “Análisis y diseño de controladores basados en lógica difusa”. Tesis de Grado para obtener el título de Ingeniero en Electrónica. Facultad de Ingeniería. Universidad de San Carlos. Guatemala.

21. Schnitman, L. y Yoneyama, T. (2001). “Takagi-Sugeno-Kang fuzzy structures in dynamic system modeling”. Proceedings of the IASTED International Conference on Control and Application pp. 160-165. Banff, Canada.

22. Trillas, E. y Gutiérrez, J. (1992). “Aplicaciones de la lógica borrosa”. Editado por Consejo Superior de Investigaciones Científicas. Madrid, España.

23. Zadeh, L. A. (1965). “Fuzzy sets: Information and control”. 8(3), pp. 338-353.

24. Zimmermann, H J. (2001). “Fuzzy Set Theory and Its Applications”. 4° ed. Boston: Kluwer Academic Publishers. ISBN: 9780792374350.

ANÁLISIS DE PROPENSIÓN EN CIENCIAS SOCIALES

PROPENSITY SCORE ANALYSIS IN SOCIAL SCIENCE

José Luis Soriano Sandoval
Dr. Carlos Fong Reynoso

Palabras clave: Propensity Score Analysis, Análisis de Propensiones, Ciencias Sociales.

INTRODUCCIÓN

El Análisis de Propensión (**PSA**. *Propensity Score Analysis*), es el método que permite solventar los problemas de heterogeneidad en el grupo de control y el grupo experimental. El método se ha utilizado desde principios de los años 90´s, principalmente en el área de la salud, no obstante, dentro de las ciencias sociales, se ha incorporado en los últimos diez años con mayor frecuencia.

De acuerdo a Pearl (2009) posiblemente el PSA es el método de emparejamiento o *Matching* más utilizado, posiblemente incluso, la estrategia más desarrollada y popular para el análisis causal en estudios observacionales. De acuerdo al autor, la herramienta se utiliza o se hace referencia en más de 127,000 artículos académicos.

De la misma manera, Hahs-Vaughn (2016) de nueva cuenta hace referencia en su libro *"Applied Multivariate Statistical Concepts"* dentro del capitulo seis, la popularidad del PSA y demuestra como se ha incorporado con mayor relevancia el uso del instrumento en los últimos años.

Lo anterior muestra de forma positiva el uso del PSA, no obstante se han identificado problemas respecto a su utilización y se han generado posturas contrarias respecto al uso del instrumento. Conforme a lo mencionado por Pan y Bai (2015), se han

descubierto tres problemas en la aplicación del método, el primero tiene que ver con la estimación del puntaje de propensión, el segundo en el análisis de resultados después del *Matching* u otro método relacionado y finalmente en el análisis de puntaje de propensión en datos complejos. Así mismo King y Nielsen (2015) afirman que el uso del PSA no debe utilizarse para realizar el *matching,* debido a que tiende a sesgar los resultados en una mayor proporción, lo que puede llegar a ser contraproducente para quien utiliza frecuentemente dicha técnica.

Con base a lo anterior, se establece un dilema respecto a la viabilidad y pertinencia que implica el uso del PSA, por lo que a continuación se exponen las posturas y los argumentos de diversos autores, referente al uso del PSA dentro de las ciencias sociales.

DESARROLLO

¿Qué es el PSA?

De acuerdo a Williams y Vogt (2011) el instrumento PSA, permite reemplaza múltiples variables de modo que solo se aplique una puntuación como predictor en lugar de múltiples variables individuales, lo que simplifica enormemente el modelo; equilibra los grupos de tratamiento y control en las variables cuando los participantes se agrupan en estratos o se su-clasifican según el puntaje de propensión; se ajusta a las diferencias a través del diseño del estudio (coincidencia) o durante la estimación del efecto del tratamiento (estratificación / regresión).

Así mismo Ho et al., (2007); Morgan y Winship, (2014) afirman que el PSA es un método cada vez más popular para procesar datos, que mejora las inferencias causales en los datos de observación. El objetivo de la correspondencia es reducir el desequilibrio en la distribución empírica de los factores de confusión previos al tratamiento entre los grupos tratados y de control. Por lo anterior se logra disminuir el desequilibrio de los grupos, y como resultado, reduce la ineficiencia y el sesgo causado por variables no relacionadas al análisis.

De igual forma, Keiffer y Lane (2016) aseguran que el PSA es un enfoque estadístico alternativo para los investigadores que buscan hacer inferencias causales utilizando grupos intactos. Un ejemplo ilustrativo de dichos autores, demostró los resultados variables del análisis de varianza, análisis de covarianza y PSA en un conjunto de datos heurísticos. Los tres enfoques se compararon por resultados y violaciones de supuestos estadísticos, los resultados demostraron cómo diferentes enfoques estadísticos pueden producir resultados variados, solo el PSA mitigó las diferencias grupales preexistentes sin violar el supuesto de independencia.

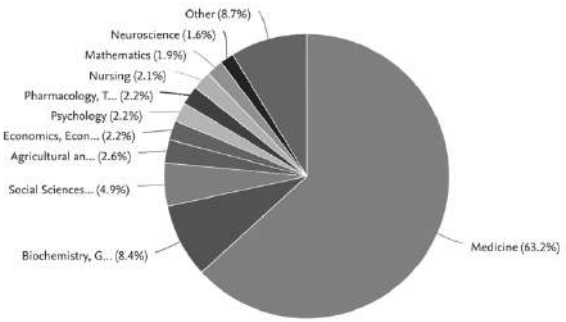
PSA en Ciencias Sociales

El PSA o Análisis de Propensiones, se ha utilizado desde hace 30 años, principalmente en el área de la salud, no obstante, dentro de las ciencias sociales el uso del instrumento es prácticamente nulo. Se comprende la existencia de mayor complejidad en el desarrollo de experimentos dentro de las ciencias sociales ya que formular experimentos que permitan controlar el mayor numero de variables, puede llegar a ser complicado, sin embargo, el PSA tiende a ser utilizado únicamente para controlar la homogeneidad de los grupos de la muestra que se pretenden contrastar, de tal manera que puede ser aplicable para el área de las ciencias sociales, como ejemplo se consideran algunos estudios donde se hace uso del PSA; "*How e-WOM and local competition drive local retailers, decisions about daily deal offerings*" de Bai, X. et al (2017), "*Consumers perceived post purchase risk in luxury services*" de Chang, Y. Ko, Y. (2017) y "*Enduring effects of goal achievement and failure within customer loyalty programs: A large-scale field experiment*" de Wang, Y, et al (2016). Las anteriores publicaciones son ejemplo del uso que se le puede dar al PSA dentro de las ciencias sociales.

A continuación, se presenta una serie de figuras que presentan de forma grafica el uso del PSA en los artículos científicos y las áreas en las que se ha utilizado dicha herramienta. Para la contabilización de las publicaciones que utilizan el PSA, se utilizo la plataforma ELSEVIER Scopus, debido a que de acuerdo

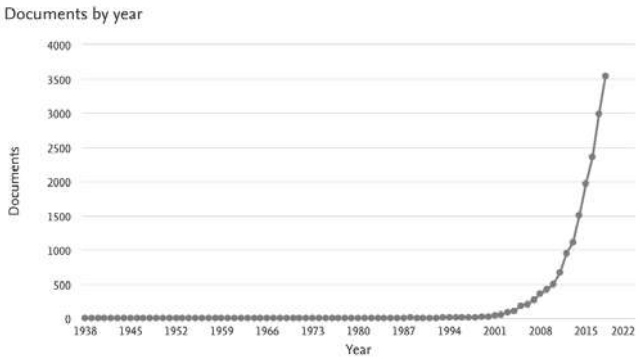
al mismo portal Scopus, es la mayor base de datos de citas y resúmenes de bibliografía revisada por pares: revistas científicas, libros y actas de conferencias, lo que permite contar con una visión más amplia de como se ha incorporado el PSA en las investigaciones científicas a lo largo de los años.

Figura 1 Distribución de investigaciones que utilizan PSA por tipo de disciplina



Fuente: Varios autores con adaptación propia.

Figura 2 Numero de publicaciones anuales que utilizan el PSA



Fuente: varios autores con adaptación propia.

DISCUSIÓN

Pertinencia y viabilidad del PSA

Conforme a lo establecido en la figura 1 y 2, se observa una tendencia positiva en el uso del PSA en los últimos diez años, no obstante, se ha generado una serie de cuestionamientos respecto a la viabilidad de las diferentes técnicas que se pueden utilizar para la aplicación del PSA.

El PSA es una técnica multivariante, porque el problema toma en cuenta a cuantas más variables se contemplan en el análisis, más se pueden homogeneizar los grupos, para lograr lo anterior, el investigador es quien debe de evaluar cuales, y cuantas variables incorpora en el PSA, con la finalidad de obtener un resultado menos sesgado.

Por lo menos existen tres distintas alternativas para realizar el Análisis de Propensiones, la primera es el *Matching*, consiste en hacer una selección entre los grupos en función del posicionamiento de las observaciones en el espacio, de tantas dimensiones como variables se contemplen en el estudio y que sean consideradas por el investigador como posibles variables que puedan generar una heterogeneidad entre grupos.

El *Matching* trata de elegir la muestra con base a proximidades y las formas para establecer dichas proximidades pueden ser mediante distintas técnicas, la más utilizada es por medio de "La distancia de Mahalanobis" sin embargo King y Nielsen (2015) proponen alternativas a dicha tecnica, debido a inconsistencias encontradas, denominándolas como la "El Dilema del *Propensity Score Matching (DPSM)"* donde concluyen que utilizar el *Matching* por medio de la distancia de *Mahalanobis*, puede llegar a ser contraproducente, debido a que puede llegar a incrementar la heterogeneidad entre los grupos cuando las distancias de las observaciones es pequeña.

De la misma manera, King y Nielsen (2015) muestran que el *Propensity Score Matching* (PSM), es un método enormemente

popular de procesamiento de datos para inferencia causal, a menudo logra lo contrario de su objetivo previsto, lo que aumenta el desequilibrio, la ineficiencia, la dependencia del modelo y el sesgo. De acuerdo a King y Nielsen (2015) la debilidad de PSM proviene de sus intentos de aproximar un experimento completamente al azar, en lugar de, como con otros métodos de emparejamiento, un experimento aleatorio completamente bloqueado más eficiente.

PSM es, por lo tanto, excepcionalmente ciego a la gran parte del desequilibrio que puede eliminarse al aproximar el bloqueo completo con otros métodos de correspondencia. Además, en datos lo suficientemente equilibrados como para aproximar la aleatorización completa, ya sea para comenzar con o después de podar algunas observaciones, PSM aproxima la coincidencia aleatoria que, como muestran en su articulo "*Why Propensity Scores Should Not Be Used for Matching"*, aumenta el desequilibrio incluso en relación con los datos originales.

King y Nielsen (2015) concuerdan en que los investigadores deben ser conscientes de que el PSM puede ayudar más en los datos donde las inferencias causales válidas son menos probables (es decir, con altos niveles de desequilibrio) y pueden causar el mayor daño en los datos que son adecuados para hacer inferencias causales (es decir, con bajos niveles de desequilibrio).

De acuerdo a Lee et al. (2010). *Mahalanobis Distance Matching* (MDM) es uno de los métodos de coincidencia más antiguos que pueden caer dentro de la clase de reducción de sesgo de igualdad de porcentaje, no obstante existen alternativas como el *Coarsened Exact Matching* (CEM), siendo parte del *Monotonic Imbalance Bounding* (MIB).

La segunda modalidad para realizar el PSA, es la estratificación y consiste en focalizar en el conjunto global de individuos, los grupos a comparar respecto a las variables que se pretende homogeneizar y crear subconjuntos en cada uno de los grupos (expuestos o no expuestos al factor de riesgo estudiado y/o gru-

po experimental) que sean homogéneos entre sí. A los grupos homogéneos les se les denomina "estratos", por lo que se mezclan los estratos y dejan fuera a los casos que no encajen en dichos estratos.

Lo anterior se puede aplicar por medio de la técnica de "Análisis clúster" y trata de elegir estratos generados a través de un Dendrograma, estratos homogéneos que tengan representantes de los dos grupos que se quieren comparar y al estar cerca en el Dendrograma, se entiende que serían propensiones similares (muestras homogéneas).

La tercera modalidad de PSA es por medio de la Regresión Logística (RL) y consiste en hacer una Regresión logística mediante variables dicotómicas: del grupo de riesgo o del grupo control y como variables independientes toman las variables que se pretende igualar. Se trata entonces, de ver cómo se comportan los coeficientes de dicha regresión y trata de ver si los coeficientes son o no estadísticamente significativos.

Un uso adicional de la RL en PSA es el siguiente: se trata de asignar, mediante la Regresión, un valor (*Propensity Score*). El *Propensity Score*, es la probabilidad de ser asignado a un grupo en función de unos valores concretos de variables independientes. Una vez que se tienen los *scores* para todos los individuos del grupo de control y grupo experimental, se procede a hacer un *Matching*, mediante la elección de individuos de ambos grupos con *scores* similares.

Las tres formas presentadas anteriormente para elaborar el PSA son alternativas desarrolladas a lo largo de los últimos 30 años, no obstante Pearl (2009) argumenta que el sesgo oculto puede en realidad aumentar debido a que igualar variables observadas puede desatar el sesgo debido a factores de confusión no observados latentes. Del mismo modo, Pearl argumenta que la reducción del sesgo sólo se puede garantizar (asintótica) modelando las relaciones de causalidad entre el tratamiento cualitativo, los resultados observados y no observados.

Es importante mencionar la influencia de la "Ley de los grandes números" dentro de los grupos de control y experimental, cuando se comparan tratamientos distintos en estudios controlados de estudios al azar, donde se eligen a los participantes de forma aleatoria, la llamada Ley de los grandes números va generando grupos que, sí el tamaño de la muestra es grande, son grupos claramente homogéneos. Por lo que adicionalmente a lo anterior, se debe de tomar en cuenta los tamaños de la muestra ya que, en muestras grandes, no será necesario el uso del PSA.

De acuerdo a Shadish (2002) una desventaja de los PSM es que sólo representa variables observadas y observables, factores que afectan a la asignación al tratamiento, pero que no pueden ser observados por lo que no pueden ser contabilizados en el procedimiento correspondiente. La confusión se produce cuando los controles experimentales no permiten al investigador, eliminar razonablemente una alternativa plausible a explicaciones de la relación observada entre las variables independientes y dependientes. La confusión se produce cuando el investigador no es capaz de controlar soluciones alternativas o explicaciones no causales para una relación observada entre las variables independientes y dependientes.

Por último, Pan y Bai (2015) afirman que desde que Rosenbaum y Rubin (1983) teorizaron el análisis de puntaje de propensión, los últimos 30 años han sido testigos de un desarrollo metodológico del análisis de puntaje de propensión que casi ha alcanzado su madurez. El análisis de puntaje de propensión se ha aplicado a muchos campos de investigación diferentes, como la medicina, la salud, la economía y la educación. Sin embargo, persisten desafíos metodológicos y prácticos para el uso del análisis de puntaje de propensión. Estos incluyen cómo evaluar la solidez del análisis de puntaje de propensión para evitar la violación de los supuestos de equilibrio, bajo qué condiciones la igualación de puntaje de propensión es eficiente, cómo implementar el análisis de puntaje de propensión de manera efectiva en datos complejos y cuáles son las consi-

deraciones relevantes después de implementar el análisis de puntaje de propensión, dichos problemas son presentados por Pan y Bai (2015) con los siguientes fundamentos.

Problemas en la estimación del puntaje de propensión

Cómo seleccionar variables es una pregunta natural en la construcción de modelos de estimación de puntaje de propensión. Intuitivamente, el investigador incluiría tantas variables observadas como sea posible en un modelo de puntaje de propensión para predecir la probabilidad de que una unidad sea asignada al grupo de tratamiento.

De acuerdo a lo mencionado por Brookhart et al. (2006) el peligro de este enfoque es que algunas variables pueden verse influenciadas por el tratamiento y por lo tanto, se viola el supuesto de ignorancia. Además, algunas variables pueden no tener ninguna asociación con el resultado, e incluir tales variables que aumentara el efecto estimado del tratamiento, mientras que el sesgo de selección no se reduce.

Además, algunos investigadores como Austin, (2011) han recomendado que los momentos de orden superior de las variables y las interacciones entre variables se examinen en los modelos de puntuación de propensión, sin embargo, Steiner y Cook (2013) mencionan que el inconveniente de estos modelos es que dependen en gran medida de los supuestos de forma funcional. En la práctica, Rubin (2001) recomendó que la selección de variables debería hacerse en base a la teoría y la investigación previa sin utilizar los resultados observados. Otra advertencia que realiza Austin (2011) para la estimación del puntaje de propensión es que el ajuste del modelo o la importancia de las variables no es de interés porque la preocupación no está en las estimaciones de los parámetros del modelo, sino más bien en el balance resultante de las variables.

Conforme a lo mencionado por Lee, Lessler y Stuart, (2010) el número de variables es grande y la forma funcional del pun-

taje de propensión parece ser compleja, algunos recomiendan utilizar métodos generalizados, modelos potenciados, un enfoque no paramétrico, de adaptación de datos, para estimar las puntuaciones de propensión.

Problemas en el análisis de resultados después del matching u otro método relacionado

Austin, (2009) relaciona que en términos de reducción del sesgo de selección, la coincidencia del puntaje de propensión suele ser mejor que la sub-clasificación, la ponderación del puntaje de propensión y el ajuste del puntaje de propensión. Por lo que se afirma que la ponderación del puntaje de propensión tiende a producir menos sesgos en las estimaciones de los efectos del tratamiento que la subclasificación. Sin embargo, la coincidencia de puntaje de propensión más la regresión con el control de variables en el análisis de resultados producirá estimaciones sólidas de los efectos del tratamiento, independientemente de la elección de los métodos de coincidencia de puntaje de propensión.

Problemas en el análisis de puntaje de propensión en datos complejos

Por último Pan y Bai (2015) argumentan que el análisis de puntaje de propensión se desarrolló originalmente en datos transversales, lo cual es común en la mayoría de los campos de investigación. A medida que los fenómenos de investigación se han vuelto multifacéticos y multidimensionales, los datos de investigación correspondientes se han vuelto cada vez más complicados, incluyendo datos longitudinales, datos multinivel y muestras de encuestas complejas. La complejidad de los datos de investigación plantea desafíos metodológicos para el desarrollo y uso del análisis de puntaje de propensión.

Cuadro comparativo de argumentos sobre el uso del PSA

Autor	Positivos	Negativos	Autor
Pearl (2009)	El PSA por medio del *Matching*, utilizando *"La distancia de Mahalanobis"*, posiblemente es la estrategia más desarrollada y popular para el análisis causal en estudios observacionales	El uso del PSA no debe utilizarse para realizar el *Matching por medio de "La distancia de Mahalanobis"*, debido a que tiende a sesgar los resultados en una mayor proporción	King y Nielsen (2015)
Hahs-Vaughn (2016)	La popularidad del PSA demuestra como se ha incorporado con mayor relevancia en los últimos años, Reduce el sesgo de grupos heterogéneos. (Todos los métodos de PSA son validos)	Una desventaja de los PSA es que sólo representa variables observadas y observables, factores que afectan a la asignación al tratamiento, pero que no pueden ser observados por lo que no pueden ser contabilizados en el procedimiento correspondiente	Sadish (2002)
Williams y Vogt (2011)	Equilibra los grupos de tratamiento y control en las variables cuando los participantes se agrupan en estratos (Se utiliza el método de Cluster)	El Peligro de este enfoque es que algunas variables pueden verse influenciadas por el tratamiento y por lo tanto, se viola el supuesto de ignorancia.	Brookhart et al. (2006)
Ho et al. (2007); Morgan y Winship, (2014)	Mejora las inferencias causales en los datos de observación. Reduce el desequilibrio en la distribución empírica de los factores de confusión previos al tratamiento entre los grupos tratados y de control.	El PSA no esta diseñado para aplicarse cuando los datos son longitudinales, multinivel o de muestras de encuestas complejas	Pan y Bai (2015)

Rubin (2001)	PSA funciona sí la selección de variables se hace en base a la teoría y la investigación previa sin utilizar los resultados observados	Todas las formas para realizar el PSA generan violaciones a los supuestos	Keiffer y Lane (2016)
Austin, (2009)	El PSA se debe de utilizar por medio de el análisis de Cluster	El inconveniente de estos modelos es que dependen en gran medida de los supuestos de forma funcional	Steiner y Cook (2013)
Lee, Lessler y Stuart, (2010)	Recomiendan utilizar métodos generalizados, modelos potenciados, un enfoque no paramétrico, de adaptación de datos, el PSA.	No funciona cuando el número de variables es grande y la forma funcional del PSA parece ser compleja	Lee, Lessler y Stuart, (2010)

Fuente: varios autores con adaptación propia.

CONCLUSIONES

Es importante tomar en cuenta los avances que se han desarrollado en los últimos 30 años para el mejoramiento del PSA y las limitaciones del propio análisis. Se logra entender que la herramienta puede generar mayores problemas sí no se aplica de forma adecuada, no obstante, para llegar a una aplicación adecuada, el investigador debe conocer a profundidad las variables que intervienen en los problemas de heterogeneidad en las muestras.

Tomando como referencia la ley de los grandes números, se puede concluir que en ocasiones, cuando las muestras tienden a ser grandes, el grupo de control y el grupo experimental, se aproximan a ser homogéneos, por lo que no sería necesario aplicar el PSA, el problema radica en la definición de "grande" debido a que en la ley de los grandes números, hace referencia a "grande" como un numero que tiende a infinito.

Por ultimo, conforme a lo señalado por, King y Nielsen (2015) el PSA no debe de realizarse por medio de PSM, debido a que no se considera un estimador estable para todas las situaciones donde el grupo experimental y el grupo de control, presentan problemas de heterogeneidad, ya que concuerdan en que los investigadores deben ser conscientes de que el PSM puede ayudar más en los datos donde las inferencias causales válidas son menos probables (es decir, con altos niveles de desequilibrio) y pueden causar el mayor daño en los datos que son adecuados para hacer inferencias causales (es decir, con bajos niveles de desequilibrio), por lo que el investigador debe de contar con criterios establecidos, que le permitan definir la pertinencia de la aplicación del PSA por medio del PSM. Es decir, establecer hasta que punto de referencia se consideran grupos homogéneos y grupos heterogéneos, con la finalidad de que el investigador pueda definir la pertinencia del uso del PSA por medio del PSM, de lo contrario buscar alternativas, por medio de la estratificación o mediante una regresión logarítmica.

REFERENCIAS

1. Austin, P. (2009) Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. *International Journal of Biostatistics,* 5(1), 1557–4679.
2. Austin, P. (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.
3. Bai, X., Marsden, J., Ross, W., Wang, G. (2017) How e-WOM and local competition drive local retailers' decisions about daily deal offerings, *Decision Support Systems*, (101), 82-94.
4. Brookhart, M. Schneeweiss, S. Rothman, K. Glynn, R. Avorn, J. y Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology, 163*(12), 1149–1156.
5. Chang, Y. y Ko, Y. (2017) Consumers perceived post purchase risk in luxury services, *International Journal of Hospitality Management, (61), 94-106.*
6. Guo, S. y Fraser, M. (2014) *Propensity Score Analysis: Statistical Methods and Applications of Advanced Quantitative Techniques in the Social Sciences*, United States of America, SAGE Publications.
7. Hahs-Vaughn, D. (2016) *Applied Multivariate Statistical Concepts*, Routledge, New York. https://doi.org/10.4324/9781315816685.
8. Ho, D. Imai, K. King, G, Stuart, E. (2007) Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, 2 (15), 199–236.
9. Iacus, M., King, G. y Porro, G. (2011) Multivariate Matching Methods that are Monotonic Imbalance Bounding, *Journal of the American Statistical Association*, 2 (106), 345–361.
10. Keiffer, G. y Lane, F. (2016) Propensity score analysis: an alternative statistical approach for HRD researchers, *European Journal of Training and Development*, 8(40), 660-675.
11. King, G. y Nielsen, R. (2015) Why Propensity Scores Should Not Be Used for Matching, *Cambridge University Press*, recuperado de: https://gking.harvard.edu/publications/why-propensity-scores-should-not-be-used-formatching

12. Lee, B. Lessler, J. y Stuart, E. (2010) Improving propensity score weighting using machine learning. *Statistics in Medicine, 29*(3), 337–346.

13. Morgan, S. y Winship, C. (2014): *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge: Cambridge University Press.

14. Pan, W. y Bai, H. (2015) *Propensity Score Analysis: Fundamentals and Developments*, New York, Guilford Press.

15. Pearl, J. (2009) The foundations of causal inference, *Sociological Methodology*, 1 (40), 75–149.

16. Pearl, J. (2009) Understanding propensity scores, *Causality: Models, Reasoning, and Inference*, Nueva York: Cambridge University Press, ISBN 978-0-521-89560-6.

17. Rosenbaum, P. y Rubin, D. (1983) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, 45*(2), 212–218.

18. Rubin, D. (2001) Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2*(3–4), 169–188.

19. Shadish, W. Cook, T. y Campbell, D. (2002) *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin. ISBN 0-395-61556-9.

20. Steiner, P. y Cook, D. (2013) Matching and propensity scores, *The Oxford handbook of quantitative methods, Oxford University Press*, 42 (1), 237–259.

21. Stuart, E. (2010) Matching Methods for Causal Inference: A Review and a Look Forward, *Statistical Science*, 1(25), 1–21.

22. Wang, Y. Lewis, C. Sprig, J. (2016) Enduring effects of goal achievement and failure within customer loyalty programs: A large-scale field experiment, *Marketing Science,* 4(35), 565-575.

23. Williams, M. y Vogt, W. (2011) *Innovation in Social Research Methods*, Londres, SAGE.

REGRESIÓN LINEAL SIMPLE, UNA TÉCNICA VIGENTE PARA LA OBTENCIÓN DE RESULTADOS EN INVESTIGACIONES CUANTITATIVAS.

Diana Corona Silva
Dr. Carlos Omar Aguilar Navarro

Palabras clave: Regresión lineal, vigencia, investigaciones cuantitativas, ventajas y desventajas.

INTRODUCCIÓN

Frecuentemente los investigadores en distintas áreas del conocimiento se interesan en conocer la forma en que en que dos o más variables se relaciona, cuestionándose si al conocer el comportamiento de una, ¿se podrá conocer el comportamiento de la otra?

Para dar respuesta a tal interrogante, de acuerdo con Mejía Trejo, (2017) examinar esa asociacion es posible mediante la tecnica de correlacion, siendo la mas sencilla de ellas la Regresion lineal simple, la cual es sujeto de discusión en el presente trabajo.

La regresion lineal simple es utilizada para estudiar la dependencia entre dos magnitudes una variable dependiente y otra variable independiente. (Montemayor Trejo, y otros, 2017).

A mas de 200 años de su primera descripcion documentada (Palacios-Cruz, 2013) a sido utlizada por diferentes disciplas, como lo son las cienias sociales, la fisica, las ciencias econimico-administrativas entre otras (Cardona Madariaga, González Rodríguez, Rivera Lozano, & Cárdenas Vallejo, 2013).

Sin embargo la regresion lineal simple es fuertemente criticada por su efectividad en la obtecion de resultados (Molnar 2019, Montero Granados 2016 y Salmerón Gómez & Rodríguez

Martínez, 2017), por lo que en el presente trabajo se discutira si la regresion lineal sigue siendo o no una tecnica de analisis multivariante con efeque dependiente vigente en dististas areas del conocimiento.

De tal manera que para el presente ensayo se tomo encuenta la base de datos Web of Science, apara analizar el uso de esta herramienta estadistica en el perido de 2013 a 2019 para destacar las áreas que mayormente la utiliza, de igual manera se presenta una tabla comparativa donde se discuten los pros y contras que ecuentran algunos autores al momento de usar esta tecnica. Se pretende que este ensayo aporte una direccion al investigador que este considerando esta herramienta para la obtencion del sus resltados.

DESARROLLO

En este apartado, de manera inicial se plantea el inicio de esta técnica en el campo de la investigación, para después definir el concepto de regresión lineal simple de acuerdo al punto de vista de distintos autores, por último se presenta un análisis bibliométrico apoyado en los datos proporcionados por la base de datos Web of Science así como una tabla comparativa en la que se discuten los pros y contras que tiene su uso.

Regresión lineal simple y su historia

Si bien la primera ocasión es que esta técnica fue descrita en el trabajo de Legendre (1805) sobre la cual se originaron las ideas modernas de la regresión, diversos autores coinciden en que la primera vez que se utilizó el termino Relación fue en el año 1886 por Francis Galton (Estepa Castro, Gea Serrano, Cañadas de la Fuente y Contreras García, 2013), (Palacios-Cruz, 2013), (Cardona Madariaga, González Rodríguez, Rivera Lozano y Cárdenas Vallejo, 2013),(Devore, 2005) (Lavalle, Micheli, & Rubio, 2006)) en su estudio "Regression towards mediocrity in hereditary stature".

A los trabajos de Galton los siguieron los de Pearson cuya gran aportacion fue señalar que era posible determinar una razón (coeficiente), cuyo valor se convierte en ± 1 cuando un cambio en cualesquiera de los órganos implica un cambio igual en el otro, y 0 cuando los dos órganos son bastante independientes (Estepa Castro, Gea Serrano, Cañadas de la Fuente, & Contreras García, 2013).

De tal manera que la información proporcionada nos habla de más de 200 años uso de esta técnica la cual será definida a continuación.

Regresión lineal simple y su definición

La regresión líneal simple ha sido definida por distintos autores de acuerdo a su punto de vista, en los siguientes párrafos se expones algunas de ellas.

De acuerdo con Bangdiwala (2018) y Moral Pelaz, (2016), la regresión lineal simple es un modelo que estudia la relación entre una sola variable dependiente Y, y una variables independientes, denotada por X.

Por su parte Szretter Noste (2017) y Carrasquilla-Batista (2016) la conciben como un modelo para el vínculo de dos variables aleatorias que se denominan X = variable predictora o covariable y la Y = variable dependiente o de respuesta. Se le denomina simple pues sólo vincula una variable predictora con Y.

Para Díaz Fernández y Llorente Marrón (2013), es la técnica que se ocupa de analizar la dependencia entre una variable dependiente y una variable explicativas.

Para (Astorga Gómez, 2014) regresión lineal simple sirve para analizar el comportamiento de las variables de entrada (o regresora) y salida (o respuesta) estableciendo predicciones y estimaciones.

Expuestas las anteriores definiciones podemos destacar tal como lo dice Cortés y Cobo (2015). La regresion lineal simple se caracterizara por conter tres elmentos, 1 variable dependiente Y , una variable independiente X y la gráfica de dispersión muestra que se relacionan por medio de una recta, cuya ecuación es y = a + bx (Cardona Madariaga, González Rodríguez, Lozano, Miller, Cárdenas Vallejo, 2013), (Mejía Trejo, 2018) y (Novales, 2010).

Analisis blibliométrico

A continuación, se hace muestra de un análisis bibliométrico utilizando como fuente primaria la base de datos Web of Science, que es una plataforma digital reconocida internacionalmente entre los investigadores por tener altos estándares de calidad (Merigó, Mas-Tur, Roig-Tierno yRibeiro-Soriano, 2015) y se ha convertido en una de las principales herramientas para la búsqueda y evaluación de diferentes tipos de publicaciones y revistas. (Gaviria-Marin, Merigó, y Baier-Fuentes, 2019).

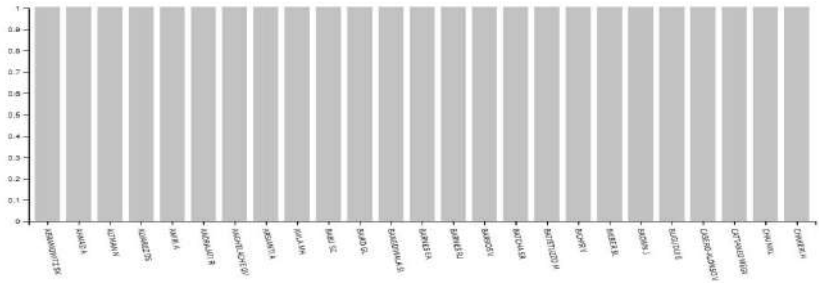
Los investigadores consideran que la Web of Science es una base de datos relevante porque proporciona un conjunto de metadatos que es esencial para este tipo de análisis, incluyendo resúmenes, referencias, número de citas, listas de autores, instituciones, países y el factor de impacto de la revista (Carvalho, 2013).

En la revisión se seleccionó un periodo de 6 años que comprende del año 2013-2019, el objetivo de ver qué autores (véase ilustración 1), en qué países se trabaja el estudio de regresión lineal (véase ilustración 2), qué tipo de documentos son publicados (véase gráfico 3), qué áreas de conocimiento estudian la técnica (véase ilustración 4) y el número de artículos publicados por año (véase ilustración 5) de tal manera que podamos dimensionar su verdadero uso.

En el lapso de 6 años la base de datos Web of Science (2019), se contabilizaron 51 registros relacionados con la Regresion

Lineal Simple, de acuerdo a los datos proporcionados no se distingue ningun autor en cuanto a produccion se refiere pues se observa que se ha publicado un articulo por investigador.

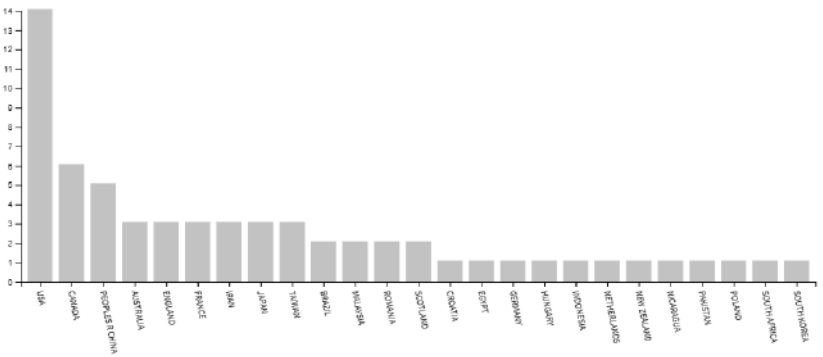
Ilustración 1 publicación por autor



Fuente: Web of Science (2019)

Sin embrago si existe una diferencia entre en la produccion de material registrado al momento de analisarla por pais de origen, siendo los estados unidos la entidad con mayor numero de divulgaciones en un periodo de 6 años.

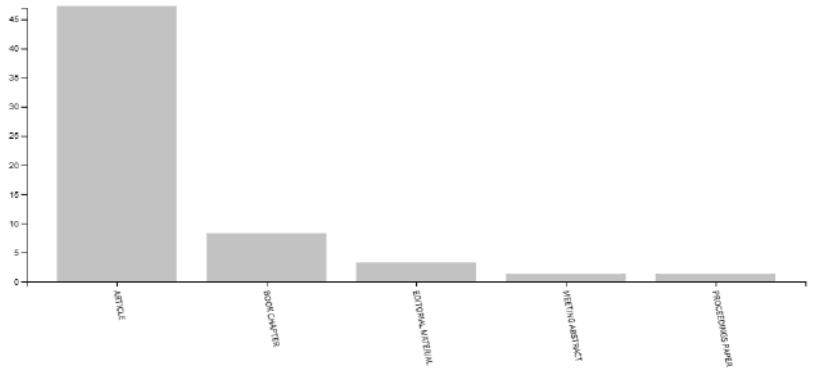
Ilustración 2 Países con publicaciones de Regresión lineal simple



Fuente: Web of Science (2019).

La aportacion al uso de la Regresión Lineal se a publicado mayormente en articulos, los cuales representan el 92% de los 51 registros en visusalizados en la base de datos.

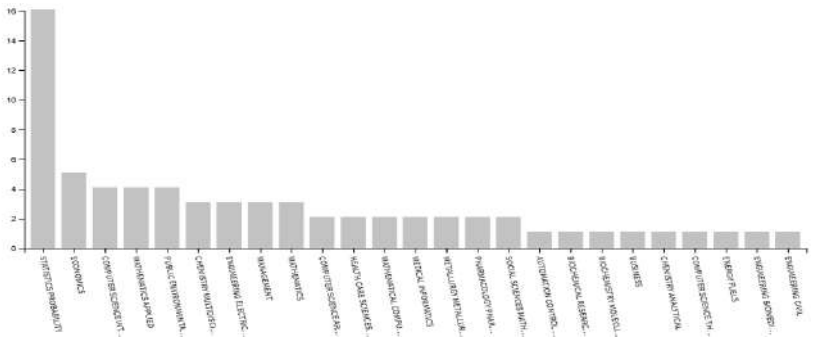
Ilustración 3 Tipo de documento



Fuente: (Web of Science, 2019)

En cuanto a las areas de conocimiento se puede observar que la probabilidad y estadisticas toman el primer lugar en produccion con 16 registros, seguida del area economica con 5 publicaciones, en tecer lugar el area de las ciencias de la computacion, las cuales representan el 31%, 9%, y 7 % respectivamente. Si bien el mayor numero de producciones es hecha por el area de probabilidad y estadistica, se puede observar que las areas que utilizan la regresion lineal simple superan la cantidad de 10.

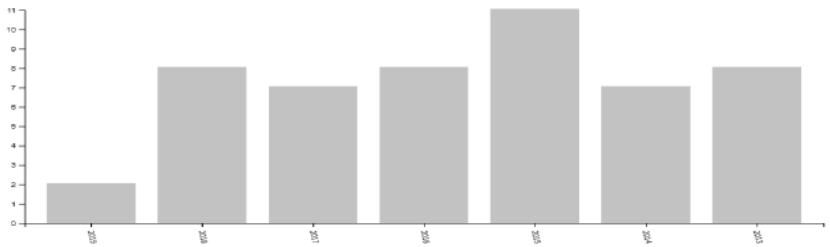
Ilustración 4 Publicación por áreas de conocimiento



Fuente: (Web of Science, 2019)

En lo referente a las publicaciones por año, se determina que el año con mayor numero de trabajos publicados en realcion a la regresión lineal simple fue el 2015, se observa una baja en los dos años siguientes, con un ligero incremento en el 2018. Es de llamar la atencion el detrimento en el año 2019 pues tan solo se registraron 2 divulgaciones.

Ilustración 5 Publicaciones por año



Fuente: (Web of Science, 2019)

DISCUSIÓN

Ya analizado su uso actual de la regresion linela simple es necesario exponer los pros y los contras que presenta el utilizar esta herramienta dentro de las investigaciones cuantitativas. Las cuales se muestran en la siguiente tabla en donde se compara los distintos puntos de vista de diferentes autores.

Cuadro comparativo de argumentos sobre el uso del Regresión lineal simple

Autor	Ventajas	desventajas	Autor
Pockels-Díaz (2012)	La revelación de información acerca de las estructuras de costos y la distinción entre los roles de las diferentes variables; además que se trata de una herramienta útil para estudiar e identificar las posibles relaciones entre los cambios observados en dos conjuntos diferentes de variables, por ultimo proporciona un medio visual para probar la fuerza de una posible relación y agilizar la toma de decisiones.	La desventaja de utilizar la regresión lineal simple es que, la misma linealidad no siempre da resultados certeros debido a que muchas cosas durante el proceso llegan a tener variaciones.	López-Briceño, (2011)

Continuación….

Molnar (2019)	En la regresión lineal simple matemáticamente, es sencillo estimar las ponderaciones además de que ofrece una predicción transparente, por lo que significa que es muchos lugares es aceptado como modelo predictivo. Una de sus grandes ventajas es la existencia de un alto nivel de experiencia pero sobre todo a la existencia de software que facilitan su implementación	La regresión lineal simple puede no tan buena en cuanto al rendimiento predictivo, porque las relaciones que se pueden aprender son tan restringidas y generalmente simplifican en exceso la complejidad de la realidad.	Molnar, (2019)
Cardona Madariaga, González Rodríguez, Lozano, Miller, y Cárdenas Vallejo, (2013)	Una ventaja de la regresión lineal simple es permite pronosticar valores futuros de la variable bajo análisis con cierto grado de certeza, lo cual constituye una herramienta poderosa pues le da al profesional la posibilidad de hacer ajustes en los procesos, tomar decisiones o establecer políticas.	Por su naturaleza, la regresión lineal simple sólo considera la relación lineales entre una variable dependiente y una independientes. Es decir, asume que existe una relación en línea recta entre ellas. A veces esto es incorrecto. Ya que en ocasiones la relación existente puede ser curva.	Flom (2019)
Cortés y Cobo (2015)	La importancia del análisis de regresión radica en el hecho de que proporciona un poderoso método estadístico que permite a una empresa examinar la relación entre dos o más variables de interés.	La regresión lineal asume que los datos son independientes. Esto significa que las puntuaciones sobre un tema no tienen relación con las de otro. Esto crea una limitación en las aplicaciones de clúster donde las variables necesitan ser agrupadas en función del espacio y el tiempo.	Salmerón Gómez y Rodríguez Martínez, (2017)

Fuente: elaboracion propia con base en varios autores.

Como se puede observar en la tabla los autores coinciden en que el uso de esta herramienta ayuda a areas economico-administrativas para hacer predicciones y de esta manera tomar deciciones.

Sin embargo otros autores postulan que el uso de esta esta tecnica no muestra la complejidad de la realidad estudiada, pues asusmen que la relacion entre las dos variables es en linea recta, restringiendo asi las relaciones que su pueden aprender.

CONCLUSIONES

La regresión lineal tiene avistamientos desde el siglo XIX y a pesar de su antigüedad se aprecia en el análisis bibliométrico que es una técnica que sigue siendo utilizada hasta la fecha. Sin embargo debido a varias de sus limitaciones ha llevado al cuestionamiento de la veracidad de los resultados que esta herramienta pueda brindar.

Al ser antigua y no haber tenido las adaptaciones necesarias esta técnica no brinda un resultado confiable de su aplicación y ese puede ser la causa del detrimento de su uso en los últimos años.

En conclusión, la utilización de la regresión lineal simple es una herramienta disponible y vigente para la investigación, sin embargo, debido a las limitaciones observadas autores como Flom (2019) recominedan el uso de otras tecnicas para resolver estas desventajas como es el uso de modelos multinivel y de esta manera aportar resultados mas fiables.

REFERENCIAS

1. Astorga Gómez, J. M. (2014). Aplicación de modelos de regresión lineal para determinar las armónicas de tensión y corriente . *Ingeniería Energética,*, 234-241.
2. Bangdiwala, S. I. (2018). Regression: simple linear. *International Journal of Injury Control and Safety Promotion*, 113-115.
3. Cardona Madariaga, D. F., González Rodríguez, J. L., Lozano, R., Miller, y Cárdenas Vallejo, E. (2013). Inferencia estadística Módulo de regresión lineal simple. *Documentos de investigación* , 45-52.
4. Cardona Madariaga, D. F., González Rodríguez, J. L., Rivera Lozano, M., & Cárdenas Vallejo, E. H. (2013). Aplicación de la regresión lineal en un problema de pobreza. *Interacción* , 74-84.
5. Carrasquilla-Batista, A., Chacón-Rodríguez, A., NúñezMontero, K., Gómez-Espinoza, O., Valverde, J., & Guerrero-Barrantes, M. (2016). Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal . *Tecnología en Marcha. Encuentro de Investigación y Extensión*, 33-45.
6. Carrollo Limeres, M. C. (2019). *Departamento de Estadística, Análisis Matemático y Optimización.* Obtenido de http://eio.usc.es/: http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat_50140116_Regr_%20simple_2011_12.pdf
7. Carvalho, M. F. (2013). An overview of the literature on technology roadmapping (TRM): contributions and trends. . *Technological Forecasting and Social Change*, 1418–1437.
8. Cortés, J. B., & Cobo, E. (2015). Regresión lineal simple. *Barcelona Universitat Politècnica de Catalunya.*, 3-35.
9. Devore, J. L. (2005). *Probabilidad y estadística para ingeniería y ciencias.* México: Thomson Learning.
10. Díaz Fernández, M., & Llorente Marrón, M. D. (2013). *Econometría.* Madrid : Ediciones Pirámide.
11. Estepa Castro, A., Gea Serrano, M. M., Cañadas de la Fuente, G. R., y Contreras García, J. M. (2013). Algunas notas históricas sobre la correlación y regresión y su uso en el aula. *Numeros , Revista de Didáctica de las Matemáticas*, 5-14.

12. Flom, P. (18 de Noviembre de 2019). *The Disadvantages of Linear Regression*. Obtenido de sciencing.com: https://sciencing.com/disadvantages-linear-regression-8562780.html.

13. Gaviria-Marin, M., Merigó, J. M., & Baier-Fuentes, H. (2019). Knowledge management: A global examination based on bibliometric analysis. *Elsevier*, 194-220.

14. Gaviria-Marin, M., Merigo, J. M., & Popa, S. (2018). Twenty years of the Journal of Knowledge: a bibliometric analysis. *Journal of knowledge Management*, 1655-1687.

15. Lavalle, A. L., Micheli, E. B., & Rubio, N. (2006). Análisis didáctico de regresión y correlación para la enseñanza media. *Revista Latinoamericana de Investigación en Matemática Educativa*, 383-406.

16. López-Briceño, E. (2011). *Análisis de regresión lineal para correlacionar datos del valor b en catálogos de sismicidad, obtenidos con dos técnicas.* Moterrey: Universidad Autónoma de Nuevo León.

17. Mejía Trejo, J. (2017). *Las ciencias de la administración y el análisis multivariante.* . Zapopan : Universidad de Guadalajara.

18. Mejía Trejo, J. (2018). *Análisis estadísitico multivariante con SPSS para las Ciencias Económico-Administrativas.* México: Cloudbook.

19. Merigó, J., Mas-Tur, A., Roig-Tierno, N., & Ribeiro-Soriano, D. (2015). A bibliometric overview of the journal of business research between 1973 and 2014. *Journal of Business Research 68* , 2645–2653.

20. Molnar, C. (2019). *Interpretable Machine Learning, A Guide for Making Black Box Models Explainable.* Leanpub.

21. Montemayor Trejo, J. A., Munguía López, J., Segura Castruita, M. Á., Yescas Coronado, P., Orozco Vidal, J. A., & Woo Reza, J. L. (2017). La regresión lineal en la evaluación de variables de ingeniería de riego agrícola y del cultivo de maíz forrajero. *Acta Universitaria*, 40-44.

22. Montero Granados, R. (2016). Modelos de regresión lineal múltiple. Documentos de Trabajo en Economía Aplicada. *Universidad de Granada*, 1-60.

23. Moral Pelaz, I. (2016). Modelos de regresión: lineal simple y regresión logística. *Revista Seden*, 195-214.

24. Novales, A. (2010). *Universidad Complutense Madrid.* Obtenido de https://www.ucm.es/: https://www.ucm.es/data/cont/docs/518-2013-11-13-Analisis%20de%20Regresion.pdf

25. Palacios Cruz, L. P.-R. (2013). Investigación clínica XVIII Del juicio clínico al modelo. *Revista Médica Instituto Mexicano del Seguro Social,*, 656-661.

26. Pockels Díaz, C. L. (17 de Diciembre de 2012). *Regresión lineal como técnica más eficiente para la previsión de la demanda*. Obtenido de https://www.eoi.es:https://www.eoi.es/blogs/scm/2012/12/17/regresion-lineal-como-tecnica-mas-eficiente-para-le-prevision-de-la-demanda/

27. Salmerón Gómez, R., & Rodríguez Martínez, E. (2017). Métodos cuantitativos para un modelo de regresión lineal con multicolinealidad. Aplicación a rendimientos de letras del tesoro. *Revista de Métodos Cuantitativos para la Economía y la Empresa,*, 169-189.

28. Szretter Noste, M. E. (2017). *Universidad de Buenos Aires.* Obtenido de http://mate.dm.uba.ar/: http://mate.dm.uba.ar/~meszre/apunte_regresion_lineal_szretter.pdf

29. Web of Science. (2019). *webofknowledge*. Obtenido de wcs.webofknowledge.com:wcs.webofknowledge.com/RA/analyze.do?product=WOS&SID=5BafQuzPfq6A5sZIAYA&field

ESPECIFICIDADES, LIMITACIONES Y PARTICULARIDADES DE LA REGRESIÓN LOGÍSTICA EN LAS CIENCIAS DE LA ADMINISTRACIÓN.

José Manuel González Gutiérrez
Dr. Antonio de Jesús Vizcaíno

Palabras claves: Especificidades, Limitaciones, Aplicación, Regresión Logística.

INTRODUCCIÓN

La regresión es una técnica estadística utilizada para analizar las relaciones funcionales entre variables (Silva y Barroso, 2004). Los modelos de regresión se utilizan para evaluar la asociación entre una variable dada, llamada variable de criterio (o dependiente en estudios experimentales), y una o más variables predictoras (o independientes). Se pueden usar con fines de estimación o predicción (Lapresa, D., et al., 2016).

Una gran cantidad de problemas de investigación requieren el análisis y la predicción de un resultado dicotómico. Éstos se han abordado tradicionalmente mediante regresión de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés) o análisis de función discriminante lineal. Ambas técnicas han revelado limitaciones para manejar resultados dicotómicos, debido a sus supuestos estadísticos estrictos (Peng et al., 2002, citado en Žáková Talpová, S. 2014, p. 2).

Existe un modelo que ha cobrado relevancia en las ciencias de la administración, hablamos de la regresión logística. De acuerdo a Žáková Talpová, S., (2014) la aplicación de esta técnica sigue siendo rara quizás como consecuencia de la literatura limitada existente en la aplicación de este método en estas ciencias.

Este trabajo pretende dar un panorama general sobre especificidades, limitaciones y particularidades del modelo de regresión logística dentro de las ciencias de la administración, además de su aportación como herramienta estadística.

DESARROLLO

La regresión logística es una de las técnicas estadístico-inferencial más utilizadas en la producción científica contemporánea, sin embargo, se tiene que potencializar en las áreas de la administración. En relación a Silva, L. y Barroso I. (2004) este modelo surge en 1960 con la aparición de un trabajo sobre el riesgo de padecer una enfermedad coronaria debido a Cornfield, Gordon y Smith (1961). De hecho, su uso o aplicación se da en mayor medida en las áreas médicas, sin embargo, esto no quiere decir que no sea efectiva en otras ramas científicas, como lo son las ciencias de la administración.

Por otro lado, la regresión logística (RL) se propuso como alternativa al método de mínimos cuadrados y al análisis de función discriminante lineal, esto fue a fines de los años sesenta y principios de los setenta. La RL se hizo disponible en paquetes estadísticos a principios de los ochenta, desde entonces su uso ha aumentado en ciencias sociales, pero se identifican pocas aplicaciones a pesar de eso.

En la década de 1980 se introdujo la regresión logística y la programación lineal, como métodos de cabecera para la construcción de puntuaciones crediticias. En la actualidad y a nivel internacional se han incorporado a esta área, técnicas de inteligencia artificial, como los sistemas expertos y las redes neuronales (Thomas, David, & Crook, 2002) por mencionar algunos avances.

El modelo de regresión logística según Ato y López (1996) utiliza una variable de criterio dicotómica y una o más variables predictivas cualitativas, ordinales o cuantitativas. La utilización del modelo logit se hace cuando queremos predecir un resultado binario, por ejemplo, "quiebra vs. no quiebra", en el caso de algún financiamiento de algún negocio y sabemos que existen varios factores que pueden incidir sobre tal resultado. Esta regresión binaria es un tipo de análisis de regresión donde la variable

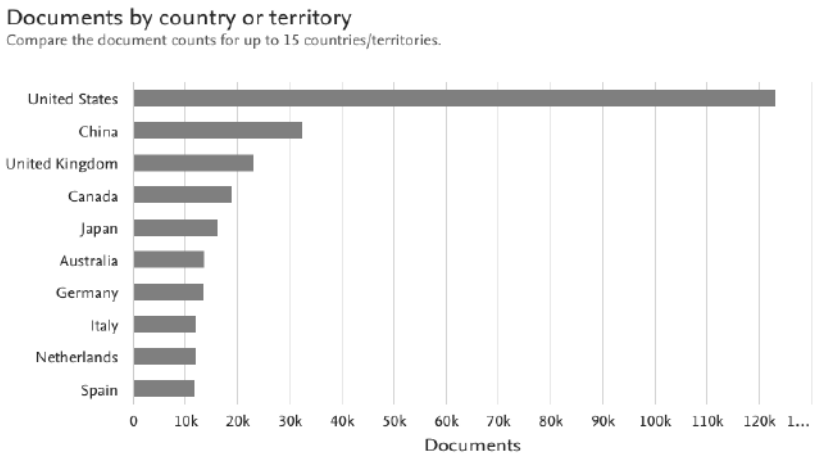
dependiente es una variable *∂ummy*[1] (variable que más adelante se explicará): código 0 (Buen Cliente) o 1 (Mal Cliente) (Castaño, F., Pérez H. y Ocaris, F. 2005, p. 58).

A continuación, se presentan algunas ilustraciones extraídas de la base de datos de Scopus, que demuestran el desarrollo y aplicación de la regresión logística a nivel mundial, por área geográfica.

Desarrollo

Como se puede apreciar, de los diez países que se encuentran, Estados Unidos es quien, en mayor medida utiliza este modelo estadístico con una producción de 122976 documentos publicados tan solo en la base de datos de Scopus y con trabajos publicados dentro del 2019, ver **Gráfico 1**.

Gráfico 1. Trabajos publicados por país

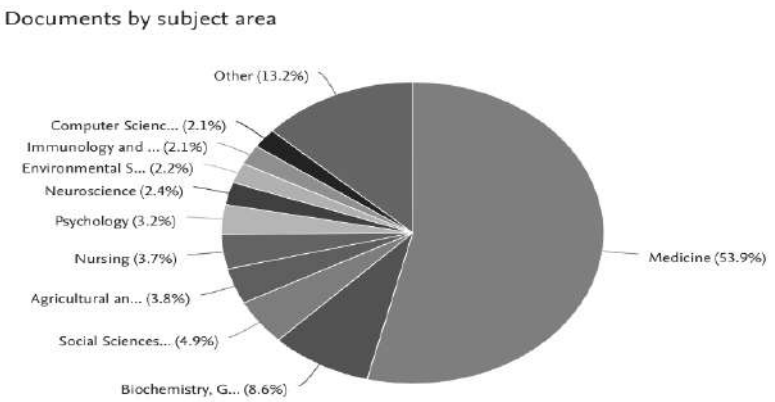


Fuente: Scopus

1.Variable Dummy: Una variable ficticia es una variable utilizada para explicar valores cualitativos en un modelo de regresión, toma un valor binario es decir valor cero o uno, ayuda a predecir cierto resultado, véase en: https://economipedia.com/definiciones/variable-ficticia.html

De igual forma se puede observar en qué área del conocimiento se han desarrollado mayormente trabajos de investigación utilizando a la regresión logística. Por ejemplo, siguiendo en la consulta dentro de la plataforma de Scopus, se aprecia en la ilustración 2., que con un 53.9% más de la mitad, en enorme desarrollo en el área médica y tan solo un 4.9% en las ciencias sociales. Ver **Gráfico 2**.

Gráfico 2. Trabajos por área



Fuente: Scopus

Uno de los problemas fundamentales en las investigaciones, cuando intervienen diversas variables en un fenómeno de estudio, es determinar cuál es la contribución de cada una de ellas, suponiendo que el resto de las variables no cambian. Por analogía, la regresión logística puede considerarse una extensión de los modelos de regresión lineal, con la particularidad de que el dominio de salida de la función está acotado al intervalo 0 y 1 y que el procedimiento de estimación, en lugar de mínimos cuadrados, utiliza el procedimiento de estimación máximo-verosímil. Es decir, a diferencia por ejemplo de la regresión lineal que, como Mejía Trejo, J. (2017, p. 192) menciona: esta se puede utilizar para analizar una única variable dependiente y varias variables independientes conociéndosele en conjunto como la ecuación de regresión, en cambio, la regresión logística utiliza variables

criterios, pero ajustándolos con datos métricos y no métricos según el ajuste de indagación que se quiera aplicar en un estudio.

En la RL es fundamental identificar como elemento diferenciador la variable dummy que, de acuerdo a Camarero, (s.f.) la categorización dummy consiste en la generación de variables dicotómicas para las distintas categorías de la variable. Estas nuevas variables se denominan *ficticias*. En la codificación binaria, 0 quiere decir que no se posee la característica y 1 que sí se posee.

Para el ajuste de modelos de regresión logística resulta esencial el empleo de programas informáticos (Camarero, L., Almazán, A. y Mañas, B., s.f.), como el SPSS o el SAS.

El modelo logit al incluir su variable dummy, permite predecir un resultado binario, al categorizar las variables seleccionadas, es decir, las codifica de manera binaria donde 0 quiere decir que no se posee la característica y 1 que sí se posee, aportando alternativa para investigaciones que requieran tener ciertos ajustes.

Existen algunas posturas de diferentes autores, en relación al potencial que tiene este modelo en comparación con la regresión lineal (múltiple o simple) aplicados a la administración, por ejemplo, siguiendo en Hurtado (2014) menciona que existen, sin embargo, muchos casos en donde la variable endógena, a pesar de que es cuantificable, no toma valores en un intervalo infinito sino sólo en un número finito de ellos. A veces esta variable ni siquiera es cuantificable como cuando se pretende caracterizar a una persona por el tipo de transporte que utiliza: 1 si usa bus 2, si usa coche particular, etc. Otros ejemplos son: decisión de tomar parte en el mercado de trabajo, el nivel de estudios de un individuo, el voto a depositar, clase de vivienda a ocupar, si se concede o no un crédito. Por ello se trata de buscar modelos que mejor se ajusten a estos casos.

Entonces se puede decir que este modelo tiene la "particularidad" para detectar datos de los casos que poseen ciertas características dentro de un estudio (Camarero, R. L., Almazán, A. y Mañas, B., s.f.).

Regresión Logística en las Ciencias del Administración

La regresión logística se usa ampliamente en la investigación biomédica (Díaz Quijano 2012). Sin embargo, en los negocios y la gestión, el empleo de la regresión logística es escaso e incluso raro visualizarlo, pero su uso va en aumento (Žáková Talpová, S., 2014).

Está técnica estadística contribuye en las ciencias sociales (Administración) para determinar cuáles variables tienen que ver con el desarrollo de un suceso, cuáles variables afectarán la toma de una decisión, o cuáles variables permitirán predecir la ocurrencia de un evento (Escobar Moreno, N.R., 2013, p. 2).

Debe tenerse en cuenta que cuando la variable dependiente es dicotómica, no pueden cumplirse los supuestos que se exigen en la regresión múltiple. Por esto, una alternativa de análisis es el análisis discriminante, pero exige normalidad multivariante de las variables independientes. Es entonces cuando puede recurrirse a la regresión logística, ésta estima directamente la probabilidad de ocurrencia de un acontecimiento, donde la variable dependiente es categórica y no continua (Escobar Moreno, N.R., 2013).

Como aportación a las ciencias de la administración la RL se ha caracterizado por suministrar un marco metodológico que permita ajustar esta herramienta de muchas formas, un ejemplo puede ser procesos de investigación de mercados aplicados en estudios desde el campo académico o empresarial y a través del SPSS. Un caso es, el de estudio de mercado que se aplicó en Colombia para fines de la utilización del modelo como muestra para académicos y aplicaciones empresariales, se puede visualizar en Escobar Moreno N (2013)[2].

Otro ejemplo puede de su utilización, fue aplicado como modelo de predicción del riesgo crediticio en las organizaciones

2. véase en: https://www.researchgate.net/publication/279981795_Analisis_de_Regresion_Logistica_para_Investigacion_de_Mercados_Logistic_Regression_Analysis_for_Marketing_Research/link/55a18fed08ae1c0e04640e77/download

de la económica social y solidaria, en un estudio aplicado por Jaime Pérez en Ecuador (2017)[3] permitió al análisis del riesgo de impago de las obligaciones crediticias que tienen las organizaciones.

DISCUSIÓN

Žáková Talpová, S. (2014) plantea una serie de estudios y aplicaciones de la RL de diferentes autores, los cuales se enlistan a continuación:

Ranyard (2012) estudió los aspectos psicológicos de las decisiones de seguro de protección de pagos. (Jayaram et al., 2010) examinaron la interrelación entre el alcance de integración de la cadena de suministro y los esfuerzos de gestión de la cadena de suministro. La regresión logística se ha utilizado para predecir posibles clientes potenciales (Akinci et al., 2007), o si las empresas van a la quiebra (Chen, 2011). También se utilizó para examinar la relación entre el proceso de socialización y el desempeño organizacional (Vinšová et.al., 2013). Uno de los temas más frecuentes en la literatura de gestión es la rotación de empleados (Tansey et al., 1996), que está representada por una variable dependiente binaria y, por lo tanto, es particularmente adecuada para la regresión logística (p.3).

Existen diferentes debates y posturas entre la regresión logística y otros métodos en la aplicación estadística; Tansey y Col. (1996) mencionaron una creciente preocupación de que el método de mínimos cuadrados se emplee cada vez más con variables dependientes binarias, violando así suposiciones importantes dentro de las investigaciones, considerando la regresión logística como una alternativa para solventar estos percances. Por su parte DeMaris (1992) señala como problema, la interpretación errónea generalizada de que la regresión logística es idéntica al análisis de regresión, lo que podría conducir a una mala interpretación del coeficiente determínante dentro las corridas estadísticas, es por eso que propone dar atención detenida dentro de su uso y

3. Véase en: https://www.uv.mx/iiesca/files/2018/03/23CA201702.pdf

aplicación. Hosmer & Lemeshow (2000) opina por su parte que la flexibilidad de la regresión logística en el análisis de conjuntos mixtos de variables nominales, ordinales e intervalos le da una gran ventaja sobre el modelado lineal, aporta más significancia al respecto y beneficia en este caso a toda ciencia donde se aplique (administración).

En relación a la inclusión de variables irrelevantes al modelo de regresión, Rao (1971) muestra que no generan sesgo en los estimadores, pero generan un incremento en sus varianzas y, por consiguiente, en sus errores cuadrados medios. Al igual que en la regresión lineal, la especificación del modelo es crucial en el análisis de regresión logística. A menudo, diferentes modelos competirán por la titularidad, por lo que los resultados pueden variar significativamente (Van der Heijden, 2012). Van der Heijden, propone una herramienta interesante que podría usarse para seleccionar el modelo de regresión logística. El proceso de selección se divide en tres etapas: selección de posibles variables independientes; construcción de diferentes modelos; y selección del modelo que se desempeña mejor de acuerdo con una medida de desempeño dada dentro del estudio. Las pruebas t simples se utilizan para seleccionar posibles variables independientes, mientras que la reducción en la probabilidad logarítmica se emplea en la selección de modelos óptimos.

El trabajo de Van der Heijden (2012) va más allá e implementa los algoritmos que se han creado en el sistema de soporte de decisiones. Este sistema tiene dos módulos: módulo de gestión del conjunto de datos y regresión logística. El módulo de gestión del conjunto de datos importa datos de una variedad de formatos (por ejemplo, MS Excel) y el módulo de regresión logística produce una coincidencia uno a uno entre sus resultados y los resultados de los paquetes comerciales (SPSS, SAS).

En las ciencias de la administración y para el desarrollo de investigaciones la regresión logística ha sido de gran uso, por ejemplo, según Hurtado Dianderas, E. (2014) esta herramienta constituye un elemento fundamental en la formación académica

de un Magíster en Administración, ya que proporciona al estudiante las herramientas necesarias para la toma de decisiones en las áreas de abastecimiento, transportes y comunicaciones en toda empresa, según lo mencionado fue aplicado en un estudio de investigación.

Por lo anterior se desprende una serie de posturas a favor y en contra de algunos autores en relación de la aplicación de la RL dentro de las ciencias de la Administración, ver **tabla n.° 1**.

Tabla 1. Posturas de "pros y contras" de la RL aplicada en áreas de la administración

Pros	**Contras**
Tansey & Col. (1996) mencionaron una creciente preocupación de que el método de mínimos cuadrados se emplee cada vez más con variables dependientes binarias, violando así suposiciones importantes dentro de las investigaciones, considerando la regresión logística como una alternativa para solventar estos percances.	DeMaris (1992) señala como problema, la interpretación errónea generalizada de que la regresión logística es idéntica al análisis de regresión, lo que podría conducir a una mala interpretación del coeficiente determínante dentro las corridas estadísticas, es por eso que propone dar atención detenida dentro de su uso y aplicación.
Hosmer & Lemeshow (2000) opina por su parte que la flexibilidad de la regresión logística en el análisis de conjuntos mixtos de variables nominales, ordinales e intervalos le da una gran ventaja sobre el modelado lineal, aporta más significancia al respecto y beneficia en este caso a toda ciencia donde se aplique (administración).	Walls y Weeks (1969) Advierten que agregar una variable a la regresión logística, nunca mejora la precisión de los estimadores de mínimos cuadrados, pero se remueve un posible sesgo y esto ocurre sin importar si la variable es importante o no.

Continuación….

Van der Heijden, propone una herramienta interesante que podría usarse para seleccionar el modelo de regresión logística.	En relación a la inclusión de variables irrelevantes al modelo de regresión, Rao (1971) muestra que no generan sesgo en los estimadores, pero generan un incremento en sus varianzas y por consiguiente en sus errores cuadrados medios.

Nota: Elaboración propia con base en: Tansey & Col. (1996), Hosmer & Lemeshow (2000), DeMaris (1992), Walls y Weeks (1969), Rao (1971).

CONCLUSIONES

La regresión logística, es una herramienta estadística que permite predecir un resultado binario, a través de una variable dummy, ésta de característica dicotómica que permite categorizar las variables seleccionadas y las codifica de manera binaria donde 0 quiere decir que no se posee la característica y 1 que sí se posee.

La regresión logística se ha utilizado en mayor medida en áreas médicas, sin embargo, en otras áreas se le ha encontrado una gran oportunidad, donde las funciones logaritmo y exponencial eliminan los problemas que ocasiona un modelo lineal (Gaméz, J., 2016, p. 4).

Siguiendo en Gaméz, J. (2016) los modelos de regresión logística se pueden considerar como un paso más de los modelos predictivos sobre conjuntos de datos. En muchos casos, o bien la dispersión de los datos, o bien su particular organización nos impedirán recurrir a los modelos de regresión lineal, pues no nos servirán para ajustar satisfactoriamente los datos y las predicciones obtenidas de ellos no serán buenas. Es por eso que este modelo brinda la oportunidad de ajustar datos a nivel estadístico, en aplicaciones de la investigación.

Como último apartado la **tabla 2** presenta algunas definiciones de la regresión logística de algunos autores consultados.

Tabla 2. Algunas definiciones de la regresión logística

Autor	**Características o definición del modelo de regresión logística**
Ato y López (1996)	Dentro del modelo logit, se utiliza una variable de criterio dicotómica, y una o más variables predictivas cualitativas, ordinales o cuantitativas.
Castaño, F., Pérez H. y Ocaris, F. (2005, p. 58.)	El modelo logit es una regresión binaria, es un tipo de análisis de regresión donde la variable dependiente es una variable dummy: código 0 (Buen Cliente) o 1 (Mal Cliente) por ejemplo.
Silva, L. y Barroso I. (2004)	La regresión logística es una de las técnicas estadístico-inferencial más utilizadas en la producción científica contemporánea. Este modelo se divide en dos: binaria múltiple y binaria simple
Camarero, L., Almazán, A. y Mañas, B. (s.f., p. 5)	El análisis de regresión logística se utiliza para estudiar la asociación entre una variable respuesta binaria con un conjunto de variables independientes. Cuando hay correlación alta entre dos variables independientes, se presentan varianzas grandes en los estimadores de los parámetros, este modelo ayuda a reducir la brecha de la no relación entre las variables.
Aldás (2011),	La regresión logística es una herramienta muy flexible para explicar la pertenencia a grupos (variable no métrica dicotómica) al permitir utilizar variables independientes métricas y no métricas.

Nota: Elaboración propia con base en: Ato y López (1996), Castaño, F., Pérez H. y Ocaris, F. 2005, p. 58. Silva, L. y Barroso I. (2004), Aldás (2011), Camarero, L., Almazán, A. y Mañas, B. (s.f., p. 5)

REFERENCIAS

1. Ato, M. y López, J.J. (1996). Análisis estadístico para datos categóricos. Madrid: Síntesis.
2. Barallobres, S. (1998). Matemática 4. Buenos Aires, Argentina: Aique.
3. Batanero, C. (2001). Didáctica de la estadística. Granada, España: Grupo de Investigación en Educación Estadística. Obtenido de la página electrónica www.ugr.es/~batanero/.
4. Batanero, C.; Godino, J. y Estepa, A. (1998). Construcción del significado de la asociación estadística mediante actividades de análisis de datos. Proceedings of the 22nd Conference of the International Group for the Psychology of mathematics Education (Vol. 1, pp. 221-236). South Africa: University of Stellenbosch.
5. Camarero, R. L., Almazán, A. y Mañas, B. (s.f.). *Regresión logística: fundamentos y Aplicación a la Investigación Sociología.* Departamento de Sociología I, UNED.
6. Castaño, F., Pérez H. y Ocaris, F. (2005). *El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito.* Revista Ingenierías Universidad de Medellín, vol. 4, núm. 6, enero-junio, 2005, pp. 55-75 Universidad de Medellín. Medellín, Colombia.
7. Carranza, O. (2004). *Logística. Mejores Prácticas en Latinoamérica.* (**2**.ª ed.). México, Thompson.
8. DeMaris, A. (1992). *Logit modeling: practical applications* Newbury Park, CA: Sage.
9. Diaz Quijano, F. A. (2012). "A simple method for estimating relative risk using logistic regression." *Bmc Medical Research Methodology*, 12.
10. Durand, A. (2006). *Logística y el Comercio Electrónico.* (**2**.ª ed.). Editorial Mc Graw Hill.
11. Escobar Moreno, N.R. (2013). Análisis de Regresión Logística para Investigación de Mercados (Logistic Regression Analysis for Marketing Research. CID, Facultad de Ciencias Económicas. Colombia, disponibel en: https://www.researchgate.net/publication/279981795_Analisis_de_Regresion_Logistica_para_Investigacion_de_Mercados_Logistic_Regression_Analysis_for_Marketing_Research/link/55a18fed08ae1c0e04640e77/download

12. Gaméz, J. (2016). *Modelización mediante regresión logística para estimación de proporciones en encuestas complejas* (Tesis de Maestría). Universidad de Granada. Granada España.

13. Gail, M. H., S. Wieand, and S. Piantadosi. 1984. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrics 71: 431–444.

14. Hocking, R. R. 1976. The analysis and selection of variables in linear regression. Biometrics 32(1):1–49.

15. Hosmer, D. W., Jr. y Lemeshow, S. (2000). *Applied logistic regression*, New York: Wiley.

16. Hurtado Dianderas, E. (2014). MODELO DE REGRESIÓN LOGÍSTICA. *Gestión En El Tercer Milenio*, *10*(20), 25 - 27. Recuperado a partir de https://revistasinvestigacion.unmsm.edu.pe/index.php/administrativas/article/view/9059

17. Lapresa, D., Arana, J., Anguera, M., Pérez-Castellaños, J., & Amatria, M. (2016). Application of logistic regression models in observational methodology: game formats in grassroots football in initiation into football. *Anales de Psicología*, 32(1), 288 294. https://dx.doi.org/10.6018/analesps.31.3.186951

18. Mejía Trejo, J. (2017). *Las ciencias de la administración y el análisis multivariante*. Edición 1, Zapopan, Jalisco, México: Universidad de Guadalajara.

19. Neuhaus, J. M. (1998). Estimation efficiency with omitted covariates in generalized linear models. J. Am. Stat. Assoc. 93(443): 1124–1129.

20. Pérez, J. (2017). LA REGRESIÓN LOGÍSTICA COMO MODELO DE PREDICCIÓN DEL RIESGO CREDITICIO EN LAS ORGANIZACIONES DE LA ECONOMÍA SOCIAL Y SOLIDARIA**.** Quito-Ecuador, Ecuador. Disponible en: https://www.uv.mx/iiesca/files/2018/03/23CA201702.pdf

21. Sifuentes Amaya, Rigoberto y Ramirez-Valverde, Gustavo. (2010). Efectos de especificar un modelo incorrecto para regresión logística, con dos variables independientes correlacionadas. *Agrociencia*, *44*(2), 197-207. Recuperado en 08 de noviembre de 2019, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-31952010000200008&lng=es&tlng=es.

22. Silva, L. y Barroso, I. (2004). *Regresión Logística*. España, Madrid: HESPÉRIDES, S. L.

23. Rao, P. (1971). Some notes on misspecification in multiple regressions. The Am. Stat. 25(5):37–39.

24. Ranyard, R. & McHugh, S. (2012). "Defusing the risk of borrowing: The psychology of payment protection insurance decisions." *Journal of Economic Psychology*, 33(4), 738-748.

25. Rosenberg, S. H., and P.S. Levy. 1972. Note: a characterization on misspecification in the general linear regression. Biometrics 28(4):1129–1133.

26. Tansey, R., White, M., Long, R. G., & Smith, M. (1996). "A comparison of loglinear modeling and logistic regression in management research." *Journal of Management*, 22(2), 339-358.

27. Thomas, L. C., David, E. B., & Crook, J. N. (2002). Credit scoring and its Applications. (S. o. Mathematics, Ed.) Philadelphia, Pennsylvania, Estados Unidos: Clarendon Press. Superintendencia de Economía Popular y Solidaria. En: http://www.seps.gob.ec/documents/20181/26626/APUNTE%20MUJERES%20SIN%20PORTADA.pdf 64d83fda-c05c-4bf9-b72e-a8c9d2278c52. Fecha de consulta: febrero de 2016

28. Van der Heijden, H. (2012). „Decision support for selecting optimal logistic regression models." *Expert Systems with Applications*, 39(10), 8573-8583.

29. Walls, R. C., and D.L. Weeks. 1969. A note on the variance of a predicted response in regression. The Am. Stat. 23(3): 24–26.

30. Žáková Talpová, S. (2012). *Strategies of Multinational and Domestic Companies in the Czech Republic.* (Ph.D. Thesis). Brno: Masaryk University.

ENFOQUES Y VALIDEZ DEL ANÁLISIS BIBLIOMÉTRICO COMO HERRAMIENTA DE REGRESIÓN LINEAL SIMPLE PARA MOSTRAR LA RELACIÓN ENTRE TURISMO Y LAS CIENCIAS ADMINISTRATIVAS.

Pilar Morales Valdez

Palabras clave: Regresión lineal simple, análisis bibliométrico, medición de producción científica.

INTRODUCCIÓN

La regresión lineal simple es una técnica de suma importancia y practicidad, gracias a su capacidad para realizar proyecciones y pronósticos de una variable dependiente explicada por una o más variables independientes (Brenes, 2018; Mejía Trejo, 2017) su campo de acción es aplicable a la investigación en múltiples áreas científica que en general tengan como objetivo pronosticar la influencia de una o más variables en sobre otras (Madariaga, González, Rivera y Cárdenas, 2013)

En la actualidad este método puede ser adaptado a cualquier investigación de carácter cuantitativo o mixto que busque predecir el comportamiento de ciertas variables a través de datos de carácter métrico que puedan facilitar su realización (Mejía Trejo, 2018). En este caso la regresión lineal se realiza a través de un análisis bibliométrico, del cual se ha hablado numerosas veces de sus funcionalidades, pero también ha sido criticado hasta la actualidad por la premisa donde se sostiene que una gran cantidad de artículos o material científico, no significa una gran cantidad de generación de conocimiento y calidad en los mismos (Sanchis Gomar, 2014; Génova, Astudillo y Fraga, 2016; Ioannidis, Boyack, Small, Sorensen y Klavans, 2014). Según esta crítica, un análisis bibliométrico solo ayudará a medir la cantidad de material y la importancia de este, en función de

su popularidad y no en función de la calidad e innovación de las aportaciones al área de conocimiento.

Para poder refutar o apoyar cualquiera de estas posturas se propone realizar un análisis bibliométrico sobre el término turismo rural. En este caso se plantea abordar al turismo desde las ciencias administrativas, en primer lugar, por la relevancia que toma al enfocarse desde las relaciones económico-sociales y la facilidad que presenta su medición a través de modelos y agentes económicos al momento de adaptar esta actividad (Streimikiene y Bilan,2015). Y en segundo lugar por la claridad que representa este como ejemplo sobre la importancia de identificar y adoptar un enfoque, al realizar una regresión lineal simple utilizando el análisis bibliométrico.

DESARROLLO

En este apartado se aborda con mayor profundidad cómo surge la técnica de regresión lineal simple y se desarrolla hasta la actualidad, a la par de uno de los métodos que se utiliza como herramienta de la misma, el análisis bibliométricos, se muestra su desarrollo, su función, y sus ventajas.

Importancia del análisis bibliométrico como herramienta de regresión lineal múltiple.

Levin y Rubin (2004) ubican el primer uso del término regresión como un concepto estadístico en el año de 1877 por Sir Francis Galton, que realizó un estudio para mostrar la tendencia de retroceso a estatura (media de la población) de los niños nacidos de padres altos. Otorgó la palabra regresión al proceso general de predecir una variable (la estatura de los niños) partiendo de otra (la estatura del padre o de la madre). Años después los investigadores estadísticos comenzarían a acuñar el término para describir el proceso por el cual se utilizan un conjunto de variables para predecir otra (Devore, 2005). A lo largo de los años ha evolucionado la definición de regresión lineal, así como ha ido creciendo su área de aplicación dentro de

las diferentes disciplinas, no obstante, Lind, Marchal, y Wathen (2015) distinguen dos elementos que se mantienen constantes: el primero, que es describir una variable y sus causalidades a través de otras. Y el segundo que busca predecir la variable dependiente, a partir de los valores de las variables independientes.

En cualquiera de los dos casos (expresado de manera estructural), la regresión lineal simple evalúa las relaciones entre una variable principal (variable dependiente, expresada en la ecuación como Y) con relación a otras variables (variables independientes, expresadas como X1, X2, X3…, Xn en la ecuación) (Moral, 2016). Lo que da significado a la información es la forma en la que se interpreta, dependiendo de lo que se quiera evaluar.

Dentro de las formas de recabar información para realizar regresiones lineales, existe una tendencia creciente por parte de los investigadores sobre el uso de Internet de las Cosas IC, es decir, de aprovechar las capacidades computacionales y de procesamiento en la nube para que les permita dar rumbo y sentido a la investigación. (Batista, Rodríguez, Núñez, Espinoza, Valverde y Guerrero, 2016).

Una herramienta que combina el método de regresión lineal múltiple con el uso del IC, es el análisis bibliométrico, este es considerado una técnica de investigación aceptada en múltiples campos como los negocios, nuevas tecnologías, elección pública o la informetría (Wagner, Roessner, Bobb, Thompson, Boyack, Keyton, Rafols y Börner, 2011).

El primer análisis bibliométrico fue realizado por Coles y Eales en 1917 cuando efectuaron un análisis estadístico de las publicaciones sobre anatomía comparativa en un lapso de tiempo de 1550 a 1860 por categorías de reino animal y distribución de países (Pérez Matos, 2002). Sin embargo, el término bibliometría no fue definido como tal sino hasta el año de 1969 como un reemplazo al término de bibliografía estadística por Alan Prichard (Ospina, 2009).

A lo largo de los años esta herramienta ha ido tomando importancia conforme el crecimiento de la base de datos como artículos, libros, reportes e investigaciones académicas que se han desarrollado y compartido a través de diferentes medios, Estrada y Cristancho (2014) Explican de manera corta el proceso de utilidad de la obtención y generación de la información: Primero la disponibilidad de información contribuye a la consolidación del proceso de búsqueda de este; segundo, el uso de la información disponible propicia el ejercicio de análisis para encontrar respuestas y soluciones a los problemas planteados; tercero, la difusión de la información generada permite la apropiación social del conocimiento y el desarrollo de la ciencia, sin dejar de lado que puede atraer visibilidad a estos investigadores y lograr la visibilidad de la información es sinónimo de reconocimiento y confirmación de los productos de investigación por parte de las comunidades científicas.

DISCUSIÓN

En este apartado se ahonda en la problemática que se observa sobre el análisis bibliométrico con respecto a la representación de los datos, se comparan las críticas y las propuestas hechas por varios autores, así como los tipos de análisis que se pueden aplicar, al final se muestra un análisis bibliométrico que ejemplifica esta problemática.

Análisis bibliométricos, tradicional e integral.

Derivado del último proceso de utilidad donde, la producción científica trae visibilidad y reconocimiento a los autores, surge una crítica que se ha mantenido a lo largo de los años relacionada con la publicación de artículos, donde la cantidad termina siendo la importante, ya que los autores publican con la intención de ser reconocidos en un sistema donde la comunidad científica basa sus decisiones en las métricas, donde los méritos de calidad quedan de lado por las consideraciones en cantidad de publicaciones de cada autor (Pulverer, 2013; Hicks, Wouters, Waltman, De Rijke, y Rafols, 2015). De esto se deriva una propuesta de indicadores

elaborada por Hicks et al. en la Conferencia Internacional sobre Indicadores de Ciencia y Tecnología que tuvo lugar el Leiden en 2014, que para fines de esta investigación se denominaran de "análisis bibliométrico integral", que en contraste con el análisis bibliométrico tradicional, manifiestan la necesidad de incluir evaluaciones cualitativas que complementen los datos numéricos que se recogen de las bases de datos, a continuación se muestra una tabla comparativa de los tipos de indicadores que representa cada propuesta de análisis bibliométrico (ver tabla 1):

Tabla 1. Comparación de análisis bibliométricos.

Análisis bibliométrico integral (basado en los indicadores de Leiden)	Análisis bibliométrico tradicional
1. La evaluación cuantitativa tiene que apoyar la valoración cualitativa por expertos 2. El desempeño debe ser medido de acuerdo con las misiones de investigación de la institución, grupo o investigador 3. La excelencia en investigación de relevancia local debe ser protegida por encima del sesgo de publicaciones anglosajonas. 4. Los procesos de recopilación y análisis de datos deben ser abiertos, transparentes y simples 5. Las diferencias en las prácticas de publicación y citación entre campos científicos deben tenerse en cuenta 6. La evaluación individual de investigadores debe basarse en la valoración cualitativa de su portafolio de investigación	1. La evaluación cuantitativa se realiza en función del objetivo de investigación del análisis. 2. El desempeño es medido en función de la cantidad de material producido. 3. El material publicado para producción internacional, tiene más relevancia que el local. 4. Los procesos de recopilación de datos, provienen de bases de datos específicas que tienen sus propias métricas. 5.No hay diferencia entre estilos de citación y publicación para los indicadores cuantitativos. 6. La evaluación de los investigadores se basa en la el número de artículos y la posición de las revistas donde publican.

Fuente: Elaboración propia

Si bien, autores como Vallejos, Torres, Sierra, León, Real (2019) y Coombes y Nicholson (2013) en sus estudios dejan en claro que esta herramienta de análisis permite identificar tendencias en la generación de conocimiento a través de la aplicación de técnicas cuantitativas que enriquecen sobre todo la revisión bibliográfica y las partes de la investigación que se deben a ella. Existe otro segmento más que pone de por medio la necesidad de validar la calidad de estas aportaciones, que no es suficiente sólo saber las tendencias y las estadísticas de publicaciones, si no las innovaciones y los avances que se hacen en cada área.

Tipos de análisis bibliométrico de material científico

Con respecto a los tipos de análisis la distinción entre indicadores es a veces difusa en el ámbito bibliométrico por la variedad de tipologías existentes derivadas de los diferentes intereses evaluativos. En materia de turismo existen múltiples investigaciones que resaltan la importancia de los estudios bibliométricos, que por medio de su metodología y sus objetivos explican de manera implícita el enfoque que presenta su análisis, con base en algunos se presenta un cuadro comparativo de los diferentes métodos de análisis bibliométrico cuantitativo (tradicional) que se pueden realizar dependiendo del tipo de información que se busque recopilar:

Tabla 2. Cuadro comparativo de tipos de análisis bibliométricos aplicados a estudios en turismo.

Tipo de análisis	Objetivo	Autores
Por colaboración entre agencias	Mostrar las redes de colaboración, entre agencias de investigación, revistas o universidades.	(Xiao, 2011; Codina-Canet, Olmeda-Gómez, y Perianes-Rodríguez, 2013)

Continuación….

Por área geográfica	Mostrar la aportación científica que se ha hecho desde un área geográfica, o para un área geográfica.	(Beckendoff, 2009; Zhong, Wu, Y Morrison, 2015)
Por tema de inserción en la disciplina de la ciencia	Mostrar cuanta importancia tiene un variable en diferentes áreas de la ciencia.	(Barrios, Borrego, y Vilaginés, 2008; Peláez-Verdet y Ferrera-Blasco, 2017)
Colaboración entre autores	Mostrar las redes de colaboración entre autores y su nivel de colaboración sobre una variable.	(Racherla y Hu, 2010; Corral-Marfil, Rodríguez, Vargas, y Cànoves, 2015)
Por autor	Mostrar la aportación y la información general sobre la aportación científica de un autor sobre una variable.	(McKener,2007)

Fuente: Elaboración propia

Las exposiciones anteriores de tipos de análisis bibliométricos en el campo del turismo pueden integrarse para determinar un análisis más completo, es decir no están limitadas por área, estas pueden complementarse dependiendo de la información que buscan obtener.

Basándonos en cuadro comparativo de tipos de análisis bibliométricos aplicados a estudios en turismo (ver tabla 2) en análisis bibliométrico integral se podría lograr si la información de la base de datos (métricos) se analizará exhaustivamente en cuanto la calidad de información que ésta ofreciera, es decir, una combinación entre cuantitativo y cualitativo, al respecto de esta integración de métodos, Tsang y Hsu (2011) por su parte proponen una clasificación de estos estudios en tres tipos:

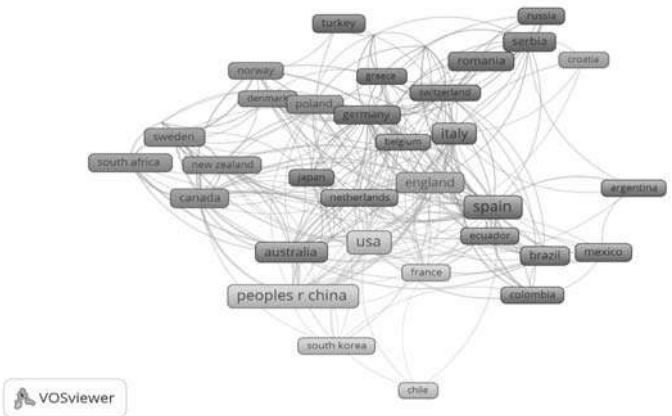
1.Ordenamiento de contribuciones de autor por rankings.

2. Análisis de las metodologías y técnicas empleadas en la investigación turística.

3. Análisis de perfil (los trabajos publicados, temas cubiertos y los lugares de publicación)

Para demostrar la funcionalidad del análisis bibliométrico como herramienta de regresión lineal y que tanto puede ser validada la calidad del material académico en turismo se presenta un análisis obtenido de la búsqueda de las palabras "Rural Tourism" en la base de datos Web of Science, con los parámetros de búsqueda delimitados a la fecha de publicación 2014-2019 (un lapso de 5 años) del cual se obtuvieron 2,052 resultados, el procesamiento de los datos para su interpretación, se muestra a través del programa "VOSviewer" (ver Gráfico 1, Gráfico 2, Gráfico 3 y Gráfico 4). Cabe mencionar que el principal objetivo de este análisis es relacionar el crecimiento de turismo rural con relación a las ciencias de la administración. Los resultados se presentan a continuación:

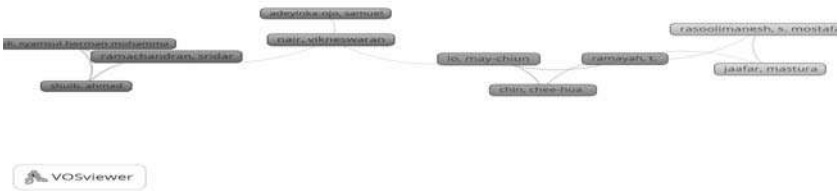
Gráfico 1. Colaboración entre países para producción académica en materia de turismo rural.



Fuente: VOSviewer a través de datos obtenidos de Web of Science.

En el gráfico 1, se puede observar el mapa que relaciona a los países que han colaborado entre sí para crear material académico sobre turismo rural, también muestra los países que más han aportado a la base de datos (los recuadros más grandes) China, España y Estados Unidos. Cada color es una red de colaboración (Clúster), en total se encuentran 5 clusters, siendo el más grande el de color rojo, que incluye 8 países en su red, y el más representativo el amarillo por su colaboración en cantidad de material científico entre dos de los tres países que más aportan. En un enfoque económico-administrativo, podría buscar comprobarse si la cantidad de producción científica tiene relación con el desarrollo económico de esta actividad, y en las redes colaborativas, que conjunto de factores económicos entre los países pudieran motivar tales colaboraciones.

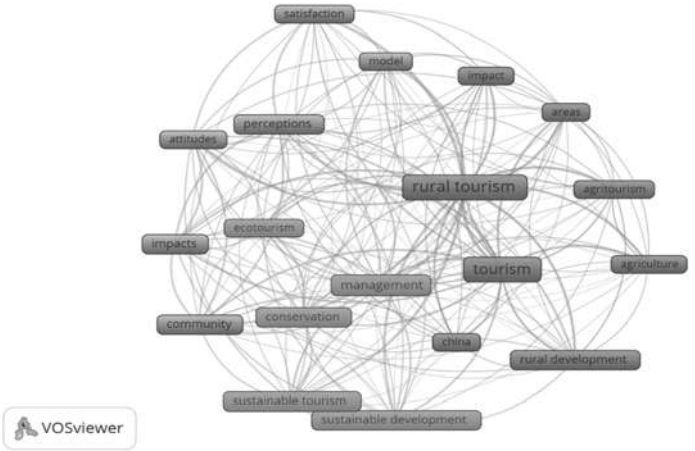
Gráfico 2. Colaboración entre autores para producción científica y académica en turismo rural.



Fuente: VOSviewer a través de datos obtenidos de Web of Science.

El Gráfico 2 representa la relación de colaboración entre autores, en relación con el número de elementos evidencia de su colaboración: mínimo 12, se eligió esta cantidad para poder mostrar las colaboraciones más significativas y posteriormente determinar bajo qué países, temas o líneas de investigación se desarrollan estas coautorías, y que intereses socioeconómicos podrían motivarlos.

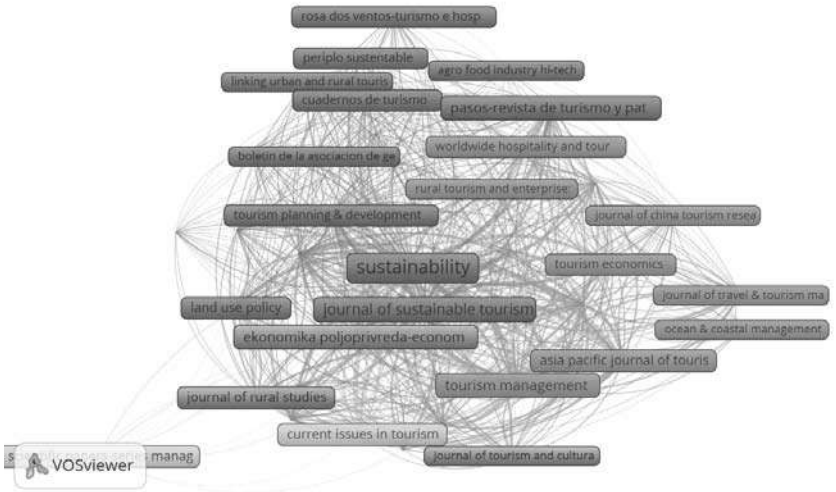
Gráfico 3. Co-ocurrencia de palabras clave en la producción científica y académica en turismo rural.



Fuente: VOSviewer a través de datos obtenidos de Web of Science.

En el Gráfico 3, se muestran las palabras clave más repetidas (utilizadas) dentro de los 2,052 resultados obtenidos de la base de datos de Web of Science, las palabras en el mapa son altamente representativas ya que tenían que aparecer como mínimo 70 veces entre las redes de investigación. Como se puede observar los clusters se dividieron en tres, siendo el rojo el conjunto más conectado con siete términos directamente relacionados, cabe destacar que administración es la segunda palabra clave más cercana a la palabra turismo rural, y la primera al término turismo. Lo cual resalta la importancia que le dan a las investigaciones turísticas desde las ciencias económico-administrativas.

Gráfico 4. Fuentes de acoplamiento de producción científica y académica en turismo rural.



Fuente: VOSviewer a través de datos obtenidos de Web of Science.

En el Gráfico 4 muestra las principales revistas y unidades de producción académica (al menos 8 documentos por unidad o revista) desde donde parte la producción en materia de turismo rural y se muestran las relaciones entre sí, se puede decir, que estos términos son la plataforma que avala y promueve la publicación de este material, así como su desarrollo. De la información que se destaca, en primera instancia sobresale el interés y nivel de especialización de la unidad en el desarrollo del tema. Por clusters se puede determinar que el más grande es el color rojo, mismo que incluye términos en español.

Finalmente, al comparar esta gráfica (4) con los resultados encontrados en la gráfica 3, se encuentra que, aunque las unidades y revistas de producción académica especializadas en administración son menos, la producción de turismo rural desde un enfoque administrativo, es bastante y no se limita solo a estas unidades.

CONCLUSIONES

Es importante relacionar la línea de investigación con el área científica a donde se desea enfocar el análisis para tener claros los objetivos sobre la información que se pretende obtener que permitirá confirmar o descartar, sobre las hipótesis e incluso el rumbo de la investigación.

En este caso la base de datos de la que se parte abarca también otras áreas científicas que cabe mencionar no se anularon para enriquecer el análisis, y aun así la puntualización en la relación que se buscaba establecer se pudo realizar, es decir, los datos obtenidos a través del análisis bibliométrico en relación con la administración mostraron una relación importante, no obstante, para enriquecer la información obtenida y darle calidad a la investigación seria necesario reforzar estos datos con un análisis contextual sobre los datos obtenidos, para finalmente aportar interpretaciones que dejan establecida la importancia de su interacción.

El ejercicio deja expuesta en este caso, la necesidad de complementar la interpretación métrica con un análisis cualitativo, para dar peso a la información obtenida, así como validez y sentido dentro de una investigación.

Si bien la literatura expuesta (ver tabla 2) nos muestra diferentes enfoques basados en los estilos de análisis bibliométricos para obtener diferente información, esto no quiere decir que se deba de utilizar solo un enfoque, siempre y cuando se tenga el claro el objetivo de la realización de regresión lineal simple, como se observa en el ejemplo, el uso de varios enfoques en suma puede contribuir a la explicación de una manera congruente y metódica, en este caso hubo dos enfoques (Ver Gráficas 3 y 4) que apoyaron directamente la relación entre estas sin mucha necesidad de ahondar en investigaciones complementarias, las dos restantes necesitarían ser enriquecidas con este tipo de análisis integral (Ver Gráficas 1 y 2). Lo que confirma la necesidad de un análisis integral.

Al final, los resultados obtenidos tanto de un análisis bibliométrico tradicional, como uno integral son válidos, en la medida en la que pueden describir datos, estimar parámetros, predecir y controlar en relación con las variables manejadas (Montgomery, Peck, & Vinning, 2002) dependiendo claro del carácter de la investigación en la que se utilicen y los objetivos de esta.

REFERENCIAS

1. Barrios, M., Borrego, A., Vilaginés, A. (2008): «A bibliometric study of psychological research on tourism». Scientometrics, (77), pp. 453-467.
2. Batista, A. C., Rodríguez, A. C., Núñez, M. K., Espinoza, G. O., Valverde, C. J., Guerrero, B. M. (2016). Regresión lineal simple y múltiple aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. Tecnología en Marcha, 30 (5), p. 33-45
3. Benckendorff, P. (2009): «Themes and trends in Australian and New Zealand tourism research: A social network analysis of citations in two leading journals (1994–2007)». Journal of Hospitality and Tourism Management, (16), pp. 1-15.
4. Brenes G, H. (2018). Aplicación del análisis de regresión lineal simple para la estimación de los precios de las acciones de Facebook, Inc. REICE: Revista Electrónica De Investigación En Ciencias Económicas, 5(10), p. 133 - 155. https://doi.org/10.5377/reice.v5i10.5535
5. Codina Canet, M.A., Olmeda-Gómez, C. y Perianes-Rodríguez, A. (2013): "Análisis de la producción científica y de la especialización temática de la Universidad Politécnica de Valencia. Scopus (2003-2010)", Revista Española de Documentación Científica, 36 (3), pp. 1-17.
6. Coombes, P., y Nicholson, J. (2013). Business models and their relationship with marketing: A systematic literature review, Industrial Marketing Management, 42(5), 656-664.
7. Corral Marfil, J.A., Rodríguez, I.M., Vargas, A. Y Cànoves, G. (2015): "Estudio de la investigación turística a través de las coautorías de artículos: cálculo de indicadores de colaboración y análisis de redes sociales. El caso de las universidades catalanas", PASOS. Revista de Turismo y Patrimonio Cultural, 13 (4), pp. 789-803.
8. Devore, J. L. (2005). Probabilidad y estadística para ingeniería y ciencias (6a ed.). México: Thomson Learning
9. Estrada, H. M., y Cristancho, M. S. (2014). The scientific information on mental health research at the University of Antioquia, Colombia. Revista Cubana de Información en Ciencias de la Salud, 25(1), pp. 4-23. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132014000100002&lng=es&tlng=en.

10. Génova G., Astudillo H,. y Fraga A. (2016) The scientometric bubble considered harmful. Sci Eng Ethics. 22(1), pp. 227-35.

11. Hicks, D,. Wouters, P,. Waltman, L., De Rijke, S., y Rafols, I. (2015) The Leiden Manifiesto for research metrics. Nature. 520. pp 429-431.

12. Ioannidis, J. P., Boyack, K. W.,Small, H., Sorensen A. A,. y Klavans R. (2014) Is your most cited work your best?. Nature. 514. Pp.561-562

13. Madariaga, C. D., González, R, L., Rivera L. M., Cárdenas V. E. (2013) Inferencia estadística: módulo de regresión lineal simple. Documentos de Investigación, 147. Retrieved from: http://www.urosario.edu.co/Administracion/documentos/Documentos-de-Investigacion/BI_147-Web.pdf

14. Mckercher, B. (2007): «A study of prolific authors in 25 tourism and hospitality journals». Journal of Hospitality and Tourism Education, (19), pp. 23-30.

15. Montgomery, D., Peck, E., y Vinning, G. (2002). Introducción al análisis de regresión lineal. Compañía editorial continental.

16. Levin, R. I. y Rubin, D. S. (2004). Estadística para administración y economía. México: Pearson Educación.

17. Lind, D., Marchal, W., y Wathen, S. (2012). *Estadística aplicada a los negocios y la economía.* (16ta ed.) México: McGraw-Hill,.

18. Mejía Trejo, J. (2017). Las ciencias de la administración y el análisis multivariante: Proyectos de investigación, análisis y discusión de resultados Tomo 1: Las técnicas independientes. Zapopan: Universidad de Guadalajara.

19. Mejía Trejo, J. (2018). Análisis estadístico multivariante con SPSS para las Ciencias Económico-Administrativas. México: Cloudbook.

20. Moral P. I. (2016) Modelos de regresión: lineal simple y regresión logística. Revista Seden. 14, pp. 195-214.

21. Ospina R, D. (2009) Caracterización de la producción científica y visibilidad de los investigadores de la Universidad Nacional de Colombia, sede Medellín, en la ISI Web of Science (1990-2007). Universidad Nacional de Colombia: Medellín.

22. Pérez Matos, N (2002). La bibliografía, bibliometría y las ciencias afines. ACIMED. 10 (1).

23. Racherla, P. Y Hu, C. (2010): «A social network perspective of tourism research collaborations». Annals of Tourism Research, (37), pp. 1012-1034.

24. Sanchis Gomar F.(2014) How does the journal impact factor affect the CV of PhD students?. EMBO reports, 15(3). pp 207

25. Streimikiene, D. y Bilan, Y. (2015). Review of rural tourism development theories. Transformations in Business & Economics, 14, 2 (35), 21-34. Obtenido de http://www.transformations.khf.vu.lt/35/ge35.pdf

26. Tsang, N.K., y Hsu, C.H. (2011): «Thirty years of research on tourism and hospitality management in china: A review and analysis of journal publications». International Journal of Hospitality Management, (30), pp. 886-896.

27. Vallejos, C. A., Torres, M. O., Sierra, M. J., León, Y. A Y Real Z. G. (2019) Identificación De Tendencias De Oferta Y Demanda Turística En El Cantón Ibarra. REVISTA INVESTIGACIÓN OPERACIONAL. 40 (4), PP. 516-522.

28. Peláez Verdet, A. y Ferrera-Blasco, M. (2017): "The usefulness of social media analysis within scholarly publications: a study of first-tier tourism journals", Tourism & Management Studies, 13 (1), pp. 43-50.

29. Pulverer, B. (2013) Impact fact-or fiction?. EMBO Publications. 32. pp 1651-1652.

30. Wagner, C., Roessner, D., Bobb, K., Thompson Klein, J., Boyack, K., Keyton, J., Rafols, I., y Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. Journal of Informetrics, 5(1), pp. 14-26.

31. Xiao, H. (2011): The capacity of a scientific community: A study of the Travel and Tourism Research Association. Journal of Hospitality & Tourism Research, (35), pp. 235-257.

32. Zhong, L., Wu, B. Y Morrison, A.M. (2015): Research on China's tourism: A 35-year review and authorship analysis, International Journal of Tourism Research, 17 (1), pp. 25-34.

EL MODELO ESTADÍSTICO DE REGRESIÓN EN CONCEPTOS CLAROS, PARA SU DIFUSIÓN Y APLICACIÓN EN EL ÁREA DE ADMINISTRACIÓN

Fecha de revisión: 19 de noviembre de 2019

Sara Guerrero Campos
Dr. Jorge Pelayo Maciel

Conceptos clave: Regresión Lineal Simple, Regresión Múltiple, Ciencias de la administración

INTRODUCCIÓN

El presente ensayo tiene como objetivo explicar los atributos del modelo estadístico de Regresión analizando paso a paso el proceso previo a su uso e interpretación. Para cumplir con el objetivo se realizó una revisión de literatura, a continuación algunas de las preguntas que serán respondidas en el ensayo ¿Qué es regresión?; ¿Para qué es útil?; ¿Cuáles son sus condiciones de aplicabilidad? y ¿Cómo se interpreta?.

La premisa del presente ensayo es explicar el modelo de forma sencilla considerando el interés en difundir las bondades de los modelos lineales entre estudiantes y profesionistas del área de las ciencias administrativas, contribuyendo con una descripción del proceso metodológico recomendado para estudios del área, fomentando que a través del análisis de datos puedan explicar, describir, correlacionar y predecir fenómenos.

DESARROLLO

El modelo de regresión es ampliamente aplicado en estudios de diversas áreas del conocimiento, sin embargo, pocos son los trabajos que explican sus componentes y describen el proceso de decisión previo a su uso y aplicación. Durante la revisión literaria pronto destacó el uso de modelos de regresión en la generación de conocimiento en las ciencias de la salud e ingenierías. En efecto, tomando como referencia su creación y su impacto, en la

Universidad de Guadalajara, las Ciencias de la Salud, Medicina e Ingenierías (RIUdeG, 2010) corresponden a las ciencias con mayor índice de producción académica, siendo las ciencias de la salud las más referenciadas en estudios posteriores, lo que se asocia con su calidad y contribución a la generación de nuevo saberes, de hecho, el modelo de regresión proviene en gran parte de estudios realizados en Biología, Biometría y Eugenesia (Estepa, Gea, Cañadas, y Contreras, 2012).

A continuación, un par de conceptos que explican ¿Qué es Regresión?:

"El análisis de regresión se usa para explicar o modelar la relación entre variable continua Y, llamada variable respuesta o dependiente y una o más variables continuas X1….Xp, llamadas explicativas o independientes" (Cayuela, 2014, p.4).

Por su parte Szretter (2017, p.29) menciona:

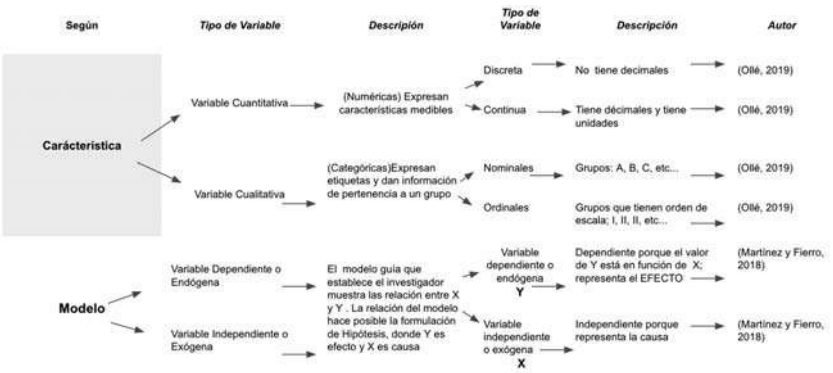
"El modelo de regresión lineal simple es un modelo para el vínculo de dos variables aleatorias que denominaremos X= variable predictora o covariable; Y= variable dependiente o de respuesta. El modelo lineal simple sólo vincula una variable predictora a Y".

Observa que en la segunda definición Szretter menciona que será simple cuando sólo se vincula la variable dependiente Y a una variable independiente X, mientras que Cayuela hace referencia a dos tipos de regresión, la simple y la múltiple, esta última aplica cuando se usan dos o más variables independientes o predictoras X relacionadas con Y. En resumen el modelo de regresión se utiliza para analizar la relación entre variables, explicar la relación entre ellas y hacer predicciones.

Analiza la diferencia en la forma en la que los autores describen las variables, Szretter le llama variable predictora o covariable a X, y a Y variable dependiente o de respuesta; mientras que Cayuela se refiere a X como variable continua independiente o

explicativa, a Y variable continua o de respuesta; para explicar el porqué, analiza el contexto, considera que los autores de artículos científicos que explican el Modelo de Regresión describen las variables con las que están trabajando, por tanto, no debe sorprenderte o confundirte si te encuentras con diferentes conceptos, deberás recordar que las diferencias en nombre corresponden al tipo de datos que se analizan y el tipo de modelo guía diseñado, antes de continuar, analiza la figura 1.

Figura 1. Tipos de variables según características y modelo



Fuente: (Martínez y Fierro, 2018; Ollé, 2019)

Ahora sabes que el modelo de regresión correlaciona una o más variables independientes X, que se define como regresión simple cuando sólo se vincula una variable independiente a Y, y regresión múltiple cuando se vinculan dos o más; además, es importante identificar el tipo y característica de la variable previo a elegir la técnica de regresión. Con esta información puedes inferir que X es independiente porque explica del fenómeno, mientras que Y es dependiente porque está relacionada con X, es decir, una variabilidad en X cambiará el valor de Y, es por tanto respuesta.

Modelo de regresión lineal simple

Se define como modelo porque su aplicación corresponde a un proceso estadístico que permite modelar la dependencia de Y con respecto a X a través de una ecuación matemática; existen varias técnicas en la categoría de modelos de regresión lineal.

A continuación se describe el tipo de técnicas de regresión lineal simple, sus condiciones de aplicabilidad, y cómo deben ser interpretados, analiza la tabla 1.

Tabla 1. Técnicas de Regresión Lineal Simple

Nombre de técnica	Condiciones requeridas	Utilidad	Interpretación
Coeficiente de correlación de Pearson (r)	Se requiere cumplir con el supuesto de linealidad, homocedasticidad, normalidad; se utiliza cuando se trabaja con datos cuantitativos, discretos y con signo positivo. Técnica sensible a datos atípicos. El modelo guía predice una relación lineal entre la variable X y Y.	Validar la propuesta de modelo, validar la probabilidad de predicción de un modelo (Lee, Yang, y Kim, 2019) Calcular el puntaje entre dos variables, para luego contrastar su variación entre sí (Mejía, 2017) además de coeficiente de correlación es útil para prueba de hipótesis	El signo de r cuando no hay relación r se acerca a 0; cuando hay relación positiva r se acerca a 1. Por lo tanto, el valor de r reside siempre entre los límites de 0 y 1 (Estepa et al., 2012).

Continuación….

Nombre de técnica	Condiciones requeridas	Utilidad	Interpretación
Coeficiente de Correlación de Spearman (rs)	Técnica de estadística analítica para correlacionar datos, se utiliza cuando se trabaja con variables cualitativas categóricos ordinales, y cuando se desconoce si existe relación lineal o si existe relación +o-.	Medida de asociación entre dos variables que no cumplen el supuesto de normalidad,a diferencia de Pearson, Spearman no es tampoco sensible a observaciones atípicas; Spearman es una técnica estadística no paramétrica (datos categóricos) analiza los datos generando rangos reemplazando cada dato observado por su rango (Szretter, 2017) además de coeficiente de correlación es útil para prueba de hipótesis.	Cuando hay una relación positiva significativa r se acerca a+1; un valor 0 indica que no hay asociación o es muy débil; y rcercano a -1indica una asociación lineal negativa.

Continuación….

Nombre de técnica	Condiciones requeridas	Utilidad	Interpretación
Coeficiente de Correlación de Kendall tau-b	Técnica de estadística analítica para correlacionar datos, se utiliza cuando se trabaja con datos cualitativos de tipo categóricos ordinales, cuando el modelo guía prevé linealidad, y existen condiciones de normalidad y de homocedasticidad.	Considerado una alternativa para la correlación de Spearman, genera una medida de asociación entre dos variables y toma en cuenta calificaciones vinculadas, puede usarse en muestras pequeñas de datos, además de coeficiente de correlación es útil para prueba de hipótesis (Mejía, 2017).	Cuando hay una relación significativa r+1; un rse acerca a 0 cuando no hay asociación o es muy débil; y cuando r -1indica una asociación con tendencia negativa.

Fuente:(Estepa et al., 2012; Lee et al., 2019; Mejía, 2017; Szretter, 2017).

En resumen, en función del tipo de variable, las características del modelo y los datos, se podrá elegir la técnica de regresión que mejor se adapte a las necesidades del investigador. La técnica de Pearson y Kendall tau -b serán opción cuando se cuente con los supuestos de linealidad, normalidad y homocedasticidad. Por linealidad, se entenderá que el modelo guía establece una relación positiva o negativa entre X y Y; normalidad hace referencia a la distribución en los datos, se dice que es normal cuando sigue una línea recta en diagonal; homocedasticidad, se refiere al supuesto de que las variables dependientes posean iguales niveles de varianza. En caso de que estos supuesto no se cumplan y se trabaje con variables ordinales (ya sea X o XY) en este caso la técnica adecuada será el Coeficiente de Correlación de Spearman.

El modelo de Regresión Lineal Simple es y=a+b.x, una variación en su expresión es y=a+bx, donde x representa la variable independiente (eje de las x abscisas); y la variable dependiente (eje de las y ordenadas) y b es la pendiente de

la recta, como se exhibe en la tabla 1. Los valores de interpretación son 1 se interpreta como una correlación fuerte con una línea en tendencia positiva; -1 correlación fuerte con una línea en tendencia negativa y 0 cuando las variables no están correlacionadas.

Modelo de regresión múltiple

El modelo de Regresión Múltiple, analiza la forma en que una variable dependiente Y, se relaciona con dos o más variables independientes, en su generalidad se utiliza p para representar la cantidad de variables independientes, la ecuación del modelo de regresión múltiple es $y=\beta 0+\beta 1x1+\beta 2x2+.....\beta pxp+\varepsilon$, donde y representa la variable dependiente, la x1,x2...xp representa las variables independientes, la β representa el coeficiente de regresión y $\beta 1,\beta 2...\beta p$ son calculados por el programa estadístico (coeficientes parciales de la regresión que miden el cambio en Y por cada cambio en X), ε el término de error, asociado a las variables no controladas también descrita como una variable aleatoria distribuida normalmente con media cero y varianza constante $\sigma 2$ para todos los valores de las X (UCA, 2019). El modelo de regresión múltiple es ampliamente aceptada y debido a su rango de aplicación es útil en la explicación y predicción de fenómenos en distintas áreas del conocimiento.

Para identificar los tipos de regresión múltiple y sus principales técnicas, analiza la tabla 2.

Tabla 2. Técnicas de Regresión Múltiple

Tipos de regresión multivariante	Descripción y tipo de técnicas
Análisis discriminante múltiple	Técnica de dependencia multivariante utilizada en modelos con variable dependiente no métricas, supuestos de aplicabilidad normalidad, homocedasticidad y linealidad. Después de calcular la función discriminante existen varios criterios estadísticos aplicables: λdeWilks;traza de Hotelling; Criterio de Pillai. Métodos para interpretar la naturaleza de los valores discriminantes métodou; Método de jack knife; F de Snedecor; análisis de asimetría Kolmogorov-Smirnov-Lillieforce KSL; comprobar homocedasticidad Test de Levene; linealidad Correlación de Pearson (Rolph, Ronald, & Hair, 2007).
Regresión logística	Técnica alternativa al análisis discriminante, aplicada cuando la variable dependiente cuente con dos categorías, existen varias técnicas estadísticas aplicables prueba ómnibus; prueba de Hosmer y Lemeshow ; R cuadrado de Nagelkerke

Fuente: (Marín, 2005); Mejía Trejo, 2017; Real, Cleries, Forné, Roso Llorach, y Martínez Sánchez, 2016).

DISCUSIÓN

Como es evidente, existen diversas técnicas estadísticas aplicables a modelos de regresión lineal simple y múltiple, cada uno corresponde al tipo de variable, modelo guía y el tamaño de los datos con el cual se aborda el estudio de un fenómeno (problema), una de las razones por las cuales son modelos ampliamente aceptados está asociado con la construcción lógica del modelo, es porque es facil de entender, por tanto, fácil de explicar y debido a los supuestos exigibles es menos propenso a un sobre ajuste.

En contrapeso, diversos autores consideran las características descritas como desventajas, dado que a medida que se complejizan los modelos de estudio, menos posibilidades se tienen de cumplir con los supuestos de normalidad, linealidad y homocedasticidad, además, el tamaño de la muestra, los datos ausentes o atípicos pueden tener una influencia sustancial en la precisión de los resultados (Mejía, 2017).

Existe la posibilidad latente de error durante el modelado del constructo, resultando en la identificación errónea de variables dependientes e independientes, restando confiabilidad y significancia a los resultados o obtenidos (Jean Pierre y Varela, 2006). Lo anterior es de interés dado que los fenómenos asociados con las ciencias administrativas suelen ser complejos y los datos de trabajo no siempre cumplen con los supuestos de aplicabilidad definidos por las técnicas de regresión simple y múltiple , por lo anterior, Técnicas de Estadística Multivariante de Segunda Generación como PLS-SEM y CB-SEM, han generado interés y popularidad (Tarka, 2018).

Otra limitante (Heizer y Render, 2007) es la dificultad para obtener registros adecuados a corto plazo en fenómenos sociales y actividades comerciales, la fidelidad o ausencia de los mismos pone en riesgo el proceso de diseño de modelo guía así como la aplicación de la técnica debido a la influencia de los datos ausentes, atípicos en la aplicación de técnicas de regresión.

A continuación las consideraciones más relevantes respecto al Modelo de regresión, ver tabla 3.

Tabla 3. Modelo de regresión, principales consideraciones

Considerando	Ventajas	Desventajas
Modelo	El modelo de regresión permite conocer si los sucesos se relacionan y con qué intensidad (Estepa et al., 2012) De aplicación multidisciplinaria; herramienta de apoyo para la toma de decisiones (Mejía, 2017) Útil para resolver problemas de identificación, para estimar coeficientes, para aceptar o rechazar hipótesis.(Stock y Trebbi, 2003) El análisis de residuos es un método efectivo para descubrir varios tipos de deficientes del modelo (Plaket, 1972)	El modelo de regresión lineal múltiple tienen menor valor predictivo en comparación con modelos de regresión no lineal simple (Baeza y Vázquez, 2014) Los errores en el modelado de constructo e identificación de variables es una posibilidad latente (Jean Pierre y Varela, 2006) El conocimiento de la situación por parte del investigador es primordial, sin el la regresión puede tener una elevada precisión predictiva sin relevancia teórica o gerencial (Mejía, 2017)

Continuación….

Considerando	Ventajas	Desventajas
Datos	El R2 ajustado corrige el R2 por el número de variables independientes del modelo, indicando la cantidad de variabilidad explicada por el modelo (Cayuela, 2014) La normalidad del término de error se puede demostrar con una gráfica de probabilidad normal de los errores residuales (Hu, 2011) En experimentos controlados, el modelo de regresión es valioso para establecer una asociación entre x, y (Hu, 2011) La técnica de Kendall tau-b, puede usarse en muestras pequeñas de datos (Mejía, 2017)	El inclumplimiento del supuesto de normalidad disminuye la eficiencia de las técnicas de regresión(Luque, 2012) La estructura de los datos y un número reducido de observaciones afectan y limitan el desempeño del modelo de regresión (Martínez, 2005) En fenómenos sociales y actividades comerciales, el acceso a información fidedigna y vigente es una limitante (Heizer y Render, 2007) Las observaciones atípicas pueden distorsionar seriamente los resultados (Plaket, 1972)

Finalmente, es evidente que algunos de los errores que se cometen al usar análisis de regresión están asociados con etapas del proceso metodológico, que a su vez resultan en la disminución de la efectividad de la técnica, por lo anterior, se recuperan las etapas del proceso metodológico propuesto por (Cayuela, 2014); Hair citado por (Mejía, 2017) y (Ollé, 2019) en primer paso definir el objetivo de la investigación, delimitar el objeto y sujeto de estudio, esbozar el modelo guía sustentado en revisión de literatura, identificar y tipificar las variables, seleccionar el método de recolección de datos, si es cuestionario, aplicar test de confiabilidad; cumplir con los supuestos de aplicabilidad, realizar la estimación y ajuste, interpretar y validar el modelo.

CONCLUSIONES

La comprensión del modelo de regresión es un requisito básico para garantizar la comprensión de numerosos conceptos y procedimientos estadísticos de primer y segunda generación, razón que motivó la elaboración del ensayo con enfoque didáctico que explica sus conceptos, características y técnicas más destacadas.

Por consiguiente, este ensayo contribuye a la difusión de técnicas de estadística aplicada para que estudiantes y profesionistas del área de ciencias administrativas las conozcan y utilicen correctamente a efecto que forme parte de su set de herramientas para fundamentar planes, proyectos y la toma de decisiones.

REFERENCIAS

1. Baeza Serrato, R., y Vázquez López, J. A. (2014). Transición de un modelo de regresión lineal múltiple predictivo, a un modelo de regresión no lineal simple explicativo con mejor nivel de predicción: Un enfoque de dinámica de sistemas. Revista Facultad de Ingeniería Universidad de Antioquia, 71, 59–71.
2. Cayuela, L. (2014). Modelos lineales: Regresión, ANOVA Y ANCOVA. Recuperado de https://portal.uah.es/portal/page/portal/epd2_asignaturas/asig202218/informacion_academica/2-Modelos%20lineales.pdf
3. Estepa Castro, A., Gea-Serrano, M. M., Cañadas, G. R., y Contreras García, J. M. (2012). Algunas notas históricas sobre la correlación y regresión y su uso en el aula. Revista de Didáctica de las Matemáticas, 81, 5–14.
4. Heizer, R., y Render, J. (2007). Administración de la Producción. México: Pearson Prentice Hall.
5. Hu, Y. (2011). Linear Regression 101. Journal of Validation Technology, 17(2), 15–22.
6. Jean Pierre, L. M., y Varela Mallou, J. (2006). Modelización con estructuras de covarianzas en ciencias sociales: Temas esenciales, avanzados y aportaciones especiales. España: Netbiblo.
7. Lee, S. E., Yang, M. S., y Kim, C. W. (2019). Evaluation of the legibility of the on-screen contents of in-vehicle transparent displays. Journal of Information Display, 20(3), 123–133. https://doi.org/10.1080/15980316.2019.1623332
8. Luque, T. (2012). Técnicas de análisis de datos en investigación de mercado (Pirámide). Madrid, España.
9. Marín, J. M. (2005). Análisis Discriminante. Recuperado de http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema6am.pdf
10. Martínez, M., y Fierro, E. (2018). Aplicación de la técnica PLS-SEM en la gestión del conocimiento: Un enfoque técnico práctico / Application of the PLS-SEM technique in Knowledge Management: a practical technical approach. RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo, 8(16), 130–164. https://doi.org/10.23913/ride.v8i16.336

11. Martínez Rodríguez, E. (2005). Errores frecuentes en la interpretación del coeficiente de determinación lineal. Anuario Jurídico y Económico Escurialense, XXXVIII, 315–332.

12. Mejía Trejo, J. (2017). Las ciencias de la administración y el análisis multivariante: Proyectos de investigación, análisis y discusión de resultados. Tomo I. Las técnicas dependientes. Zapopan, Jalisco, México: Universidad de Guadalajara.

13. Ollé, J. (2019). 6 Técnicas Estadísticas Indispensables. Recuperado de https://conceptosclaros.com/que-es-regresion-logistica/

14. Plaket, R. L. (1972). Studies in the history of the probability and Statistics XXIX. The discovery of the method of least squares. Biometrika, 59, 239–251.

15. Real, J., Cleries, R., Forné, C., Roso-Llorach, A., y Martínez Sánchez, J. M. (2016). Utilización de los modelos de regresión múltiple en estudios observacionales (1970-2013) y requerimiento de la guía STROBE en revistas científicas españolas. SEMERGEN - Medicina de Familia, 42(8), 523–529. https://doi.org/10.1016/j.semerg.2015.06.020

16. RIUdeG. (2010). Producción Científica. Recuperado de Producción Científica Universidad de Guadalajara Red Universitaria de Jalisco website: https://riudg.udg.mx/

17. Stock, J. H., y Trebbi, F. (2003). Retrospectives: Who invented instrumental variable regression? The Journal of Economic Perspectives, 17(3), 177–194.

18. Szretter, M. E. (2017). Apunte de Regresión Lineal. Recuperado de http://mate.dm.uba.ar/~meszre/apunte_regresion_lineal_szretter.pdf

19. Tarka, P. (2018). An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. Quality & Quantity, 52(1), 313–354. https://doi.org/10.1007/s11135-017-0469-8

20. UCA. (2019, noviembre 8). Regresión y Correlación: Fórmulas básicas en la regresión lineal simple [Educativa]. Recuperado el 8 de noviembre de 2019, de Universidad Central de Arkansas website: http://www.uca.edu.sv/matematica/upload_w/file/REGRESION%20SIMPLE%20Y%20MULTIPLE.pdf

EL ANÁLISIS DE DATOS USADO ENTRE REGRESIÓN LOGÍSTICA Y/O REGRESIÓN LINEAL EN ESTUDIOS DE DISCAPACIDAD

Emanuel Vicuña Huerta
Dr. Gabriel Salvador Fregoso Jasso

Palabras Clave: Estudios de Discapacidad, Regresión Lineal, Regresión Logística.

INTRODUCCIÓN

¿Qué es la regresión lineal y la regresión logística? ¿Para qué sirven? ¿Qué método de análisis de datos predomina entre regresión logística y/o regresión lineal en estudios de discapacidad? Estas son algunas de las preguntas que resolveremos con este ensayo.

El ensayo tiene como objetivo identificar cuál regresión es más utilizada en estudios de discapacidad y así descubrir cuál es la tendencia que tienen estos estudios.

También hay que considerar que estas herramientas son muy útiles para la predicción estadística, claro cada una con sus diferencias en cuanto lo que se consigue de predicción.

Este escrito se dividirá de la siguiente manera en la primera parte veremos a ver lo que es la regresión en la logística, la segunda parte vamos a ver la regresión lineal, en la tercer parte cómo emplean estudios de discapacidad en diversas áreas, ya sea la regresión logística y la regresión lineal, y ya por último la cuarta parte vamos a realizar la discusión y conclusión de este ensayo.

DESARROLLO

En esta parte del ensayo el autor intenta explicar que son las regresiones lineales, y la regresión logística.

Asimismo, hace un análisis bibliométrico para ver la posible tendencia en el uso de estas dos regresiones y las áreas que se usan más para estudios de discapacidad.

De igual manera se hace una tabla con conjunto de diferentes estudios que se realizaron al usar estas técnicas para analizar datos. Con ejemplos y comparación en su uso.

Regresión logística

La técnica de la regresión logística se inicia en los años 60 con el trabajo de Cornfield, Gordon y Smith (1961). Walter y Duncan (1967) ya la utilizan en la forma actual, siendo usada a partir de los años 80 gracias al avance tecnológico con que se cuenta desde entonces. (López-Roldán y Fachelli, 2015).

En general, la regresión logística es conveniente cuando la variable de respuesta es politómica (varias categorías de respuesta), pero es principalmente útil cuando solo hay dos posibles respuestas (variable de respuesta dicotómica), que es el caso más usual (Fernández, 2011).

La regresión logística mezcla dos prácticas del análisis estadístico: el análisis de tablas de contingencia con el tratamiento de modelos log-lineales, y el análisis de regresión por mínimos cuadrados ordinarios (López-Roldán y Fachelli, 2015).

En ambos asuntos nos hallamos con prohibiciones que la regresión logística soluciona: en el primer asunto los modelos de dependencia no podían utilizar variables continuas y en el segundo las variables categóricas no siempre funcionan como buenos predictores. (López-Roldán y Fachelli, 2015).

Al considerar al análisis de regresión logística como técnica destinada al análisis de una relación de dependencia, nos referiremos fundamentalmente a ella como una técnica predictiva y, no tanto como técnica destinada a establecer relaciones de causalidad, si bien implícitamente se razone la causalidad. (López-Roldán y Fachelli, 2015).

La regresión lineal va a contestar a preguntas tales como: ¿Se puede predecir con antelación si un cliente que solicita un préstamo a un banco va a ser un cliente moroso? ¿Se puede predecir si una empresa va a entrar en bancarrota? ¿Se puede predecir de antemano que un paciente corra riesgo de un infarto? (Fernández, 2011).

El análisis de regresión logística posee dos particularidades: la regresión logística binaria cuando se intenta comprender una peculiaridad o acontecimiento dicotómico (estar desempleado o no, abstenerse en las elecciones o no), y la regresión logística multinomial en el tema más usual de pretender exponer una variable cualitativa politómica. Para ello se requiere convertir la variable en diversas variables dicotómicas ficticias, es decir, creando tantas variables dicotómicas (*∂ummy*).(López-Roldán y Fachelli, 2015).

VARIABLES DUMMY: Las variables aclaratorias de tipo nominal con más de dos categorías deben ser incluidas en el modelo definiendo variables *∂ummy*. (Fernández, 2011).

Variables cualitativas en el modelo logístico: Como la metodología disponible para la estimación del modelo logístico se basa en la utilización de variables cuantitativas, al igual que en cualquier otra manera de regresión, es inconveniente que en él interactúen variables cualitativas, ya sean nominales u ordinales. La asignación de un número a cada categoría no resuelve el problema. La solución a este problema es designar tantas variables dicotómicas como número de respuestas. Estas nuevas variables, simuladamente creadas, reciben en la literatura anglosajona el nombre de dummy, traduciéndose con diferentes denominaciones como pueden ser variables internas, indicadoras, o variables diseño (Fernández, 2011).

Regresión logística en diferentes tipos de estudios

Estudios Descriptivos:

La regresión logística puede aprovecharse como técnica descriptiva cuando se desea estudiar desde una perspectiva habitual a la aparición de un determinado evento en un grupo de individuos (Universidad Carlos III de Madrid).

Análisis de Factores de Riesgo:

La regresión logística puede utilizarse como técnica para la estimación de la razón de disparidad (odds ratio *OR*) (Universidad Carlos III de Madrid).

Evaluación de la Interacción:

Pondremos dos factores de exhibición (variables dicotómicas) conseguimos definir el conflicto para los distintos niveles de exposición y calcular el *OR* para cada uno de estos rangos (Universidad Carlos III de Madrid).

Análisis de regresión logística binaria simple

La regresión logística binaria se identifica por acomodar una variable dependiente cualitativa con dos valores (categorías o grupos) que conforman la existencia y el abandono de un rasgo explícito. Por ejemplo, las personas que estudian y las que no, las que compran en días de descuento y las que no, las personas que ven un partido de futbol y las que no, los ciudadanos que se abstienen en las elecciones y los que no, los que votan a un partido y los que no, los consumidores que compran un producto y los que no, las personas que están en paro y las que no, etc. (López-Roldán y Fachelli, 2015)

Los rasgos específicos por la variable dependiente se intentan explicar en función de una cadena de variables independientes o predictoras que nos establecen en qué se distinguen los dos

conjuntos. Si reflexionamos tan sólo una variable independiente podemos hablar de regresión logística simple, si consideramos dos o más variables independientes el modelo de regresión logística es múltiple. (López-Roldán y Fachelli, 2015)

El modelo de regresión logística binaria toma dos acontecimientos de un fenómeno o variable, precisos y absolutos, que se catalogan con valores 0 y 1. Si la probabilidad de que ocurra uno de ellos y, la otra de que la probabilidad de que la otro ocurra es igual a 1 menos la posibilidad. (López-Roldán y Fachelli, 2015)

La idea es tomar en cuenta los datos de una (o más variables en la adaptación múltiple) para precisar un modelo que pueda pronosticar la probabilidad de la variable dependiente, es decir, se trata de hallar una o más variables que excluyan bien entre los dos posibles valores de la variable. (López-Roldán y Fachelli, 2015)

Condiciones de aplicación

Se establecen las siguientes condiciones. (López-Roldán y Fachelli, 2015) :

a. El modelo debe estar exactamente especificado y ser relevante.
b. No se excluyen variables independientes distinguidas.
c. Las variables independientes se calculan sin equivocaciones.
d. Las observaciones son autónomas entre sí.
e. Ausencia de colinealidad entre las variables independientes. Es una cuestión de grado. Correlaciones de 0,8 la implican pues incrementa los errores típicos: cuando los errores típicos sean superiores a 2 indica la existencia de multicolinealidad. La colinealidad se puede detectar, pero no es fácil de resolver. Cuando es alta o no tolerable, hay que revisar el modelo y, por ejemplo, eliminar una o más

de las variables colineales, cambiar la escala de medida o combinarlas en una medida única.

f. Linealidad de las variables cuantitativas.

g. Monotonicidad: cada independiente interactúa de forma directa o indirecta (López-Roldán y Fachelli, 2015).

h. En relación al tamaño de la muestra. Hosmer y Lemeshow recomiendan muestras mayores de 400 casos. De Maris (1992) sugiere 15 casos por variable (López-Roldán y Fachelli, 2015).

Regresión lineal.

Antes de entrar al tema de la regresión de lleno vamos a señalar los diferentes tipos de variables existentes

Nominales: estas son variables que se consideran categóricas sin ninguna en las que se alinen jerárquica, como son las dicotómicas en se señala si el suceso es positivo o negativo, aquí un ejemplo se graduó o no el estudiante, también otra variable nominal es la de procedencia. Dicha variable se puede ordenar alfabéticamente, etc. (Reding Bernal, Zamora Macorra, y López Alvarenga, 2011)

Ordinales: aquí la variable adquiere valores categóricos que tienen una ordenanza jerárquica, un ejemplo es, las variables que registran los niveles de alguna enfermedad. En cada uno de estos se puede calcular en orden de los estadios. (Reding Bernal, Zamora Macorra, y López Alvarenga, 2011)

|Continuas: la variable puede ser numérica, de manera positiva o negativa, de presión arterial, días, etc. (Reding Bernal, Zamora Macorra, & López Alvarenga, 2011)

Regresión lineal simple

La regresión lineal simple es útil para encontrar la potencia o capacidad de cómo se corresponden dos variables: una independiente, que se representa con una X, y otra dependiente, que se

identifica con una Y; sin embargo, la regresión lineal simple ella se puede diferencia de otros métodos, pues con ella puede evaluar o pronosticar la validez de la variable de respuesta a partir de un valor dado a la variable explicativa. (Reding Bernal, Zamora Macorra, y López Alvarenga, 2011)

Supuestos del modelo de regresión lineal simple

Normalidad: las equivocaciones tienen un repartimiento normal con media de cero y con variancia constante. Esto pretende indicar que los valores de Y siguen una colocación normal. Cuando este supuesto no se compensa, antes de realizar un modelo de regresión podría realizarse una transformación de la variable Y, en la que la nueva variable se disperse cerca en escritura normal (Reding Bernal, Zamora Macorra, y López Alvarenga, 2011).

Independencia: son las variables independientes; es decir, el error en una variable no depende de otra variable, lo que significa que la variable Y, la variable es independientes. Esto puede ser modificado cuando se hace observación de Diseños longitudinales o Diseños de tendencia y de evolución de grupo (Reding Bernal, Zamora Macorra, y López Alvarenga, 2011).

Homocedasticidad (homogeneidad de la variancia): aquí nos indica que la variabilidad del error es constante y es la misma para todos los errores y como consecuencia la variancia de Y es la misma para diferentes valores fijos de X (Reding Bernal, Zamora Macorra, y López Alvarenga, 2011).

Linealidad: indica que, cuando ya se tienen los valores fijos de X, los valores de Y forman una línea recta. Esta teoría se representa $Y/X = \beta 0 + \beta 1X$, donde $\beta 0$ es la dinámica entre los valores promedio de la variable Y cuando la variable explicativa X vale cero. Cuando los valores de la variable explicativa analizados no incluyen al cero, la interpretación de $\beta 0$ no tiene sentido. $\beta 1$ es la pendiente de la recta. (Reding Bernal, Zamora Macorra, y López Alvarenga, 2011).

Regresión lineal múltiple

La regresión Lineal Múltiple nos aprueba establecer la conexión que se produce entre una variable dependiente y el grupo de variables independientes (X1, X2, ... XK). El análisis de regresión lineal múltiple, se distingue del simple, se aproxima más a situaciones de análisis real puesto que los fenómenos, hechos y procesos sociales, por definición, son complejos y, en consecuencia, deben ser explicados en la medida de lo posible por la serie de variables que, directa e indirectamente, participan en su concreción (Rodríguez-Jaume y Mora Catalá, 2001).

Al usar el análisis de regresión múltiple lo más común es que la variable dependiente como las independientes sean medidas en escalas de razón o intervalo. Pero también pueden ver otras situaciones donde usaremos este análisis en variables dependientes continuas con variables categóricas o igualmente se aplica el análisis de regresión lineal múltiple en situaciones de relación de variable dependiente nominal con un conjunto de variables continuas (Rodríguez-Jaume y Mora Catalá, 2001).

En el análisis de regresión múltiple, ensayos y análisis que se aplican para establecer la dependencia y grado de asociación entre una variable dependiente y sus supuestas variables aclaratorias, así como los rangos de los parámetros de la ecuación, no difieren de los determinados en el análisis de regresión simple (Rodríguez-Jaume y Mora Catalá, 2001).

En el análisis de regresión múltiple se necesitan cálculos estadísticos más laboriosos, por lo contrario, pruebas y análisis, al contrario, el análisis de regresión lineal simple en el análisis con la relación de un par de variables el proceso se resolvía en un solo paso. (Rodríguez-Jaume y Mora Catalá, 2001).

El análisis de regresión lineal múltiple se hace por varios pasos aquí el siguiente procedimiento implica que:

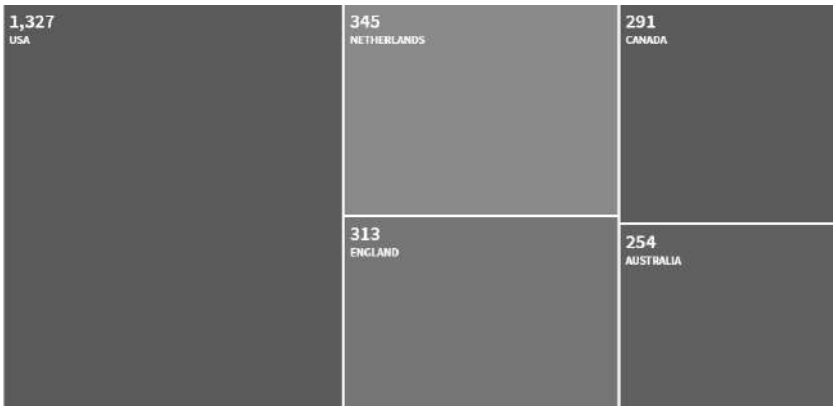
1. Usar variables con criterios necesarios (Rodríguez-Jaume y Mora Catalá, 2001).

2. Valorar si las variables siguen con los criterios necesarios (Rodríguez-Jaume y Mora Catalá, 2001).
3. Si es necesario los ajustes de los datos al modelo de regresión lineal. (Rodríguez-Jaume y Mora Catalá, 2001).

El uso de la regresión lineal y/o regresión logística en estudios de discapacidad

A continuación, vamos a ver varios ejemplos donde se usan las dos técnicas de regresión que acabamos de explicar. Todos los ejemplos que se pondrán a continuación tienen que ver con el tema de discapacidad, también se verá un poco de cada investigación con la finalidad de saber cómo se empleó el tipo de regresión y por qué. Pero antes de ver los ejemplos vamos a ver un análisis bibliométrico para saber de qué tipo de regresión se usa mas en los estudios de discapacidad y área.

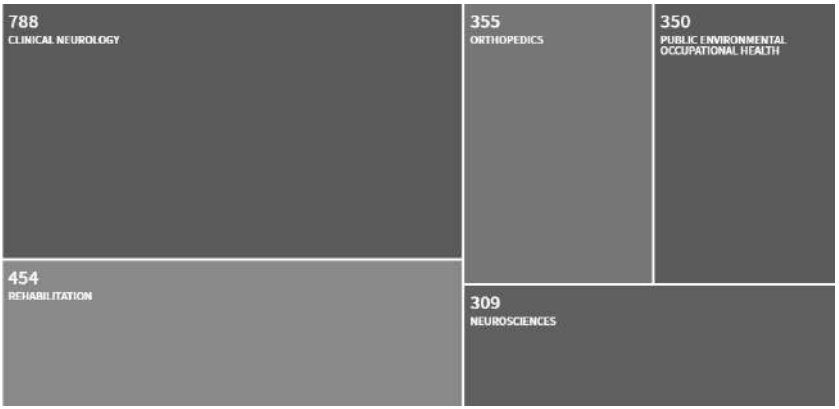
Figura 1. Los países con más publicaciones en temas de discapacidad y uso de la regresión lineal.



Fuente: Web of Science. 2019

Análisis bibliométrico de los 5 países con más publicaciones en cuestión de uso de regresión lineal en estudios de discapacidad de un total de 3445 documentos con estas dos variables, este análisis se realizó en Web of Science.

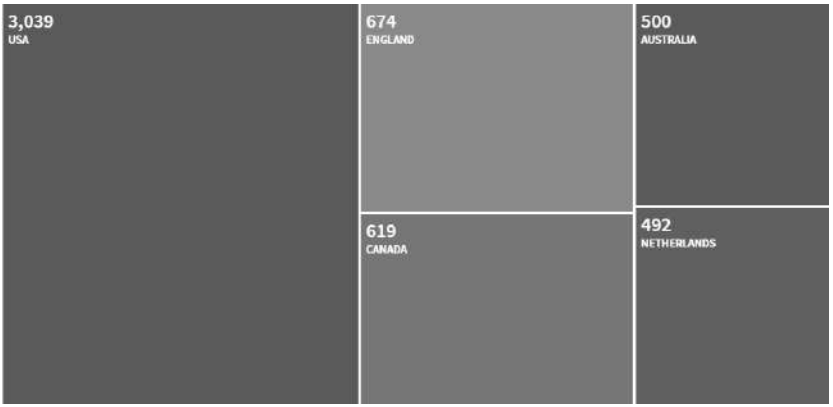
Figura 2. Los temas con más publicaciones en temas de discapacidad y uso de la regresión lineal.



Fuente: Web of Science. 2019

Análisis bibliométrico de las 5 áreas con más publicaciones en cuestión de uso de regresión lineal en estudios de discapacidad de un total de 3445 documentos con estas dos variables, este análisis se realizó en Web of Science.

Figura 3. Los países con más publicaciones en temas de discapacidad y uso de la regresión logística.



Fuente: Web of Science. 2019

Análisis bibliométrico de los 5 países con más publicaciones en cuestión de uso de regresión logística en estudios de discapacidad de un total de 8099 documentos con estas dos variables, este análisis se realizó en Web of Science.

Figura 4. Los temas con más publicaciones en temas de discapacidad y uso de la regresión logística.

Fuente: Web of Science. 2019

Análisis bibliométrico de los 5 áreas con más publicaciones en cuestión de uso de regresión logística en estudios de discapacidad de un total de 8099 documentos con estas dos variables, este análisis se realizó en Web of Science.

Tabla 1 – El uso de la regresión lineal y/o regresión logística en estudios de discapacidad

Objetivo	Análisis de datos	Resultados
"Se plantea estimar la relación entre pobreza extrema y discapacidad en Colombia, basado en el Censo General de 2005." (Herazo Beltrán y Domínguez Anaya, 2013).	"Los datos obtenidos de las diferentes fuentes de datos se analizaron mediante modelos lineales de regresión." (Herazo Beltrán y Domínguez Anaya, 2013).	"El parámetro que refleja la correlación entre las dos variables se conoce como coeficiente de correlación (r). Este parámetro estima la fuerza de la asociación entre estas dos variables continuas. Del presente análisis, se pudo obtener en concreto un coeficiente de correlación positivo, el cual indica que, al aumentar el porcentaje de pobreza, aumenta el porcentaje de discapacidad en un departamento; o un coeficiente de correlación negativo, indicando que, si el valor de pobreza extrema aumenta, el porcentaje de discapacidad disminuye." (Herazo Beltrán y Domínguez Anaya, 2013).

Continuación….

Objetivo	Análisis de datos	Resultados
"Analizar la relación entre el exceso de peso y la condición de discapacidad en las personas mayores de la Argentina y evaluar en qué medida podría estar operando algún factor protector que reduzca o atenúe el efecto del exceso de peso sobre la pérdida de capacidades funcionales en las personas mayores de 64 años" (Monteverde, 2015).	"Los datos obtenidos de las diferentes fuentes de datos se analizaron mediante regresión logística" (Monteverde, 2015).	"Se concluyó que la mayoría de las personas mayores de 64 años de edad tendrían al menos una de las tres condiciones de discapacidad consideradas para dicho estudio" (Monteverde, 2015).
"Identificar la percepción que poseen profesorado y alumnado sobre la calidad de vida de los estudiantes con discapacidad de centros de formación laboral, y más concretamente, si existe relación entre la percepción del profesorado y la percepción del alumnado" (Castro, Casas, Sánchez, Vallejos, y Zúñiga, 2016)-	"Escala Objetiva y la Escala Subjetiva. En segundo lugar, se han efectuado 5 modelos de regresión lineal múltiple mediante el método de pasos sucesivos para comprobar la predicción de las dimensiones evaluadas en la escala subjetiva sobre la percepción del profesorado, evaluando las dimensiones de la calidad de vida de los estudiantes "(Castro, Casas, Sánchez, Vallejos, y Zúñiga, 2016).	"Parece que las percepciones de los profesionales sobre si una persona con discapacidad es autodeterminada se refieren a que la propia persona con discapacidad sea consciente de su propia autodeterminación. En cuanto a la inclusión social, los resultados han mostrado que las percepciones de los profesionales sobre esta dimensión tienen como único predictor la propia percepción de la persona con discapacidad". (Castro, Casas, Sánchez, Vallejos, y Zúñiga, 2016)

Continuación….

Objetivo	Análisis de datos	Resultados
"Se analiza la variabilidad de la discapacidad por comunidades autónomas desde una doble vertiente, los factores individuales y del entorno" (Gispert Magarolas, et al., 2009).	"Se realizó una regresión logística de dos niveles. El primer nivel correspondió a los individuos y el segundo a la comunidad autónoma" (Gispert Magarolas, et al., 2009).	"Las características individuales no explican suficientemente la variabilidad de la discapacidad entre CCAA y no se han identificado variables del entorno que sean significativas" (Gispert Magarolas, et al., 2009).
"Establecer un modelo predictivo del grado de discapacidad en adultos con lesión medular a partir de la utilización del WHO-DAS II". (Henao Lema & Pérez Parra, 2011).	"Se construyó un modelo de regresión lineal múltiple para discapacidad. " (Henao Lema & Pérez Parra, 2011).	"El mejor modelo predictivo de discapacidad en adultos con lesión medular con más de seis meses de evolución se construyó con las variables tiempo de evolución, índice sensitivo ASIA y desempleo por la lesión. " (Henao Lema & Pérez Parra, 2011).
"Se estimo el impacto de tener discapacidades sobre la probabilidad de estar laboralmente activo en Colombia, a partir de la Encuesta de Calidad de Vida 2013, desde un enfoque de género. "(Acuña, 2016).	"Se estimaron diferentes modelos de regresión logística que incluyen, además de las características sociodemográficas, nueve diferentes tipos de discapacidad permanente." (Acuña, 2016).	"Los resultados más importantes, destacan una alta correlación positiva, para ambos sexos, entre los niveles de formación y la participación laboral, así como el impacto negativo de las limitaciones de movilidad y habla, en el caso de las mujeres, y de los problemas mentales o de aprendizaje en el de los hombres. "(Acuña, 2016).

Continuación….

Objetivo	Análisis de datos	Resultados
"Determinar qué factores clínicos predicen la discapacidad y la mala calidad de vida en pacientes con trastorno de ansiedad generalizada." (Marian López de la Parra, Mendieta Cabrera, Muñoz Suarez, Díaz Anzaldúa y Cortés Sotres, 2014).	"Se trató de un estudio descriptivo y transversal. Los datos se analizaron mediante frecuencias, porcentajes y promedios. Se utilizó un análisis de regresión lineal para determinar cómo predicen los factores clínicos y demográficos la discapacidad y la mala calidad de vida." (Marian López de la Parra, Mendieta Cabrera, Muñoz Suarez, Díaz Anzaldúa y Cortés Sotres, 2014).	"Encontramos que la presencia de antecedentes heredofamiliares de trastornos de ansiedad, así como mayores puntajes en la Escala de Depresión de Hamilton, predicen una menor calidad de vida, a diferencia de una mayor edad, la cual predice una mayor calidad de vida en estos pacientes. Mayores niveles de discapacidad se asociaron con el sexo masculino, una menor edad de los pacientes, comorbilidad con trastornos del Eje II, presencia de antecedentes heredofamiliares de trastornos de ansiedad y mayores puntajes en la Escala de Depresión de Hamilton." (Marian López de la Parra, Mendieta Cabrera, Muñoz Suarez, Díaz Anzaldúa y Cortés Sotres, 2014).

Continuación….

Objetivo	Análisis de datos	Resultados
"Determinar las principales variables asociadas a discapacidad funcional en pacientes con EA." (Marengo, Schneeberger, Gagliardi, Maldonado Cocco, & Citera, 2020).	"El análisis estadístico, correlación de Pearson de las principales variables. Las variables continuas fueron comparadas por test de Student y ANOVA. Las posibles variables asociadas a discapacidad funcional fueron analizadas por regresión lineal." (Marengo, Schneeberger, Gagliardi, Maldonado Cocco, & Citera, 2020).	"La actividad de la enfermedad fue la principal variable asociada a discapacidad funcional en pacientes con EA, justificando un 60% de las variaciones del BASFI."(Marengo,Schneeberger, Gagliardi, Maldonado Cocco, & Citera, 2020).

Fuente: Análisis de diversos estudios de discapacidad usando de la regresión lineal y/o regresión logística.

DISCUSIÓN

Usando los indicadores y la tabla anterior podemos deducir que el tipo de regresión que más se usa es la logística, incluso más que la regresión lineal, ¿a que se debe esto?, vamos hacer una tabla comparativa donde se discutirán las diferencias entre ambas regresiones.

Tabla 2 – Comparación de la regresión lineal y regresión logística en requerimientos del modelo

Requerimientos/Hipótesis del modelo de regresión	**Requerimientos en el modelo de regresión logística**
La utilización inferencial plena del modelo de regresión requiere: 1. Para cada conjunto fijo de x la distribución de y debe ser normal con Media. 2. La varianza de y es constante para cualquier valor de x 3. Las observaciones de y son independientes entre sí. 4. El número de variables explicativas es menor que el de observaciones. (Nolasco, 2016)	• A diferencia del modelo de regresión lineal, las inferencias no necesitarán suposición en repartición alguna. Se dirá que las inferencias son asintóticas, es decir, válidas para un muestra suficientemente grande (Nolasco, 2016).

Fuente: Estadística avanzada en ciencias de la salud: modelos lineales con adaptación propia.

Tabla 3 – Comparación de la regresión lineal y regresión logística en construcción del modelo

Construcción de un modelo de regresión lineal múltiple. Etapas	**Construcción de un modelo de regresión logística. Etapas**
Etapa 1: Especificación de variables y modelo propuesto. Etapa 2: Estimación del modelo. Etapa 3.- Validación de la hipótesis de linealidad. Bondad de ajuste del Modelo. Etapa 4.- Verificación de requerimientos. Análisis de residuos. Etapa 5.- Inferencias con el modelo.	Etapa 1: Especificación de variables y modelo propuesto. Etapa 2: Estimación del modelo. Etapa 3: Bondad de ajuste del modelo. Etapa 4.- Inferencias con el modelo.

Fuente: Estadística avanzada en ciencias de la salud: modelos lineales con adaptación propia.

Con qué podemos concluir y hacer el cierre de la discusión, que la regresión logística se usa mas debido a que la mayoría de los estudios que se realizan con el tema de discapacidad son descriptivos, y muy poco correlacionales, transaccionales, etc.

CONCLUSIONES

Para terminar este ensayo vamos a hacer lo por puntos, oraciones concretas y simples.

- La mayoría de los estudios de discapacidad son en el área de la salud, cuando ya podemos realizar más estudios en las aéreas económicas-administrativas.
- En general los estudios de discapacidad son descriptivos porque seguimos desconociendo mucho de ese tema.
- En los estudios que son de discapacidad se usa la regresión logística, por lo debido a que son estudios descriptivos y no a profundidad.
- Como punto de vista del autor, las regresiones no necesariamente se tienen que usar una u otra, sino pueden usar una de la otra de manera sino que pueden complementarse, para dejar un aporte científico más basto y profundo.

REFERENCIAS

1. Acuña, Ó. A. (2016). Participación laboral de personas en situación de discapacidad. Análisis desde un enfoque de género para Colombia. *Economía: Teoría y Práctica*, 137-167.
2. Castro, L., Casas, J. A., Sánchez, S., Vallejos, V., y Zúñiga, D. (2016). Percepción de la calidad de vida en personas con discapacidad. *Estudios Pedagógicos XLII*, 39-49.
3. Fernández, S. d. (2011). *Regresión Logística.* Madrird.
4. Gispert Magarolas, R., Clot-Razquin, G., March Llanes, J., Freitas Ramírez, A., Busquets Bou, E., Ruíz-Ramos, M., y Rivero Fernández, A. (2009). PREVALENCIA DE LA DISCAPACIDAD EN ESPAÑA POR COMUNIDADES. *Rev Esp Salud Pública*, 821-834.
5. Henao Lema, C. P., y Pérez Parra, J. E. (2011). Modelo predictivo del grado de discapacidad en adultos con lesiones medulares: resultados desde el WHO-DAS II. *Rev. Cienc. Salud.*, 159-172.
6. Herazo Beltrán, Y., y Domínguez Anaya, R. (2013). Correlación entre Pobreza Extrema y Discapacidad en los Departamentos de Colombia. *Ciencia e Innovación en Salud*, 11-17.
7. López-Roldán, P., y Fachelli, S. (2015). *Metodología de la investigación social cuantitativa.* Barcelona · España.
8. Madrid, U. C. (s.f.). *Introducción a la regresión logística.*
9. Marengo, M., Schneeberger, E., Gagliardi, S., Maldonado Cocco, J., y Citera, G. (2020). Determinantes de discapacidad funcional en pacientes con espondilitis anquilosante en Argentina. *Revista Argentina de Reumatología*.
10. Marjan López de la Parra, M., Mendieta Cabrera, D., Muñoz Suarez, M. A., Díaz Anzaldúa,, A., Y Cortés Sotres, J. F. (2014). Calidad de vida y discapacidad en el trastorno de ansiedad generalizada. *Salud Mental*, 509-516.
11. Monteverde, M. (2015). Exceso de peso y discapacidad en las personas mayores de la Argentina. *Salud Colectiva*, 509-521.

12. Nolasco, A. (2016). ESTADÍSTICA AVANZADA EN CIENCIAS DE LA SALUD. *Este documento ha sido publicado en el Repositorio de la Universidad de Alicante*.

13. Reding Bernal, A., Zamora Macorra, M., Y López Alvarenga, J. C. (2011). *¿Cómo y cuándo realizar un análisis de regresión lineal simple?* México, DF.

14. Rodríguez-Jaume, M.-J., Y Mora Catalá, R. (2001). *Análisis de regresión múltiple.* Universidad de Alicante. Servicio de Publicaciones.

Se terminó de imprimir en *Junio 2020*
en los Talleres Gráficos de
Prometeo Editores, S.A de. C.V.
Libertad 1457, Col. Americana,
C.P. 44160, Guadalajara, Jalisco

La edición consta de 100 ejemplares
Impreso en México / Printed in Mexico

La presente obra, *Ensayos 2019. Análisis Multivariante con Enfoque Dependiente en las Ciencias de la Administración como base para la Innovación*, pretende reunir una serie de ensayos elaborados por los estudiantes del Doctorado de Ciencias de la Administración (DCA) del Centro Universitario de Ciencias Económico Administrativas (CUCEA) de la Universidad de Guadalajara (UdeG), basados en lo aprendido en la asignatura de Investigación Cuantitativa I. Dichos ensayos, se orientan en principio a realizar un ejercicio de disertación que refuerce ya sea la argumentación de su tesis en la parte metodológica o bien, sea una contribución a la materia. Para ambos casos se resalta la pertinencia de su redacción a partir de la introducción para desarrollar los conceptos y/o modelos que justifican la base de los puntos antagónicos a tratar siendo la base para realizar la discusión que permite aclarar la contribución esperada. Finalmente, se exponen los puntos de conclusión esenciales que sirvan al lector y al expositor, para estudios posteriores.

CUCEA

ISBN: 978-607-98782-6-9