

Editorial

La Estadística como pilar fundamental de la Analítica de Datos

Statistics as a fundamental pillar of Data Analytics

Juan Luis Arias Vargas, M.Sc.

La estadística ha sido parte de la historia del hombre desde tiempos inmemorables, pues este siempre ha tenido el interés de cuantificar y tener información de sus pertenencias, es así como de una forma muy rudimentaria se realizaba estadística para tener control a nivel personal de las posesiones y a nivel estatal del número de personas, familias y parcelas, con el fin de poder recaudar los impuestos, pero sólo hasta los siglos XVII y XVIII [1], es cuando la estadística se convierte en una ciencia con los aportes de Graunt quien realizó el primer censo moderno y creó la primera tabla de frecuencias por edades, pero fue Achenwall quien acuñó la palabra Estadística, sentando de esta forma las bases de la estadística descriptiva. Por otro lado, Fermat, Pascal, Bernoulli, Laplace, Gauss, Poisson, Bayes y Markov, entre otros, fueron los precursores de la teoría de la probabilidad y de la incorporación de esta en la estadística [2], dando origen a todo el desarrollo de la estadística moderna que comprende lo relacionado con Distribuciones de Probabilidad, la Inferencia Estadística, el Muestreo Estadístico, el Diseño de Experimentos, los Modelos de Regresión, Series de Tiempo y Análisis Multivariado.

Es así como la estadística se ha convertido en una de las ciencias más importantes en la era moderna de la humanidad, pues el papel de esta es fundamental para tomar decisiones de forma acertada, bien sea con toda la información (datos poblacionales) o con parte de la información (datos muestrales) sobre las variables de una población o fenómeno de estudio.

Lo anterior aplica a nivel personal, empresarial y gubernamental. En este sentido, actualmente no es posible pensar que una persona tome una decisión sobre la compra de un bien sin antes realizar una indagación de modelos, precios, pros, contras y otras variables de interés, tampoco se podría pensar que una empresa tome decisiones sobre lanzar un nuevo producto al mercado o desee introducirse en un nuevo mercado, sin haber realizado un estudio estadístico riguroso que le permita medir el riesgo que está asumiendo, ni mucho menos considerar que un gobierno tome decisiones de inversión, subsidios y en general de cualquier índole, sin antes tener claro el comportamiento de las variables de aquellos entornos que se van a afectar, en todos estos casos es necesario el uso de herramientas estadísticas.

No obstante, la ciencia y la tecnología crecen a pasos agigantados y la estadística no es ajena a este fenómeno, en este sentido, esta ha evolucionado y se ha adaptado a las nuevas necesidades personales, sociales, empresariales y gubernamentales, que deben tomar decisiones de forma mucho más rápida y eficaz teniendo en cuenta que cada día se produce mucha información (datos estructurados y no estructurados) que proviene principalmente de la información propia de cada empresa o gobierno, de sensores utilizados para este fin y de la web a partir de todas las aplicaciones diseñadas para capturar la información de interés.

En congruencia con lo anterior, toda esa información debe ser organizada y utilizada de la mejor forma posible para la toma de decisiones, y de esto se encarga la Analítica de Datos, la cual tiene como pilar fundamental la Estadística, que gracias a su desarrollo como ciencia, en sinergia con los avances computacionales y el desarrollo de software han permitido la generación de sendos y robustos modelos matemáticos como: la Minería de Datos (Data Mining) usada para el análisis exploratorio de datos [3]; Regresión Logística, Árboles de Decisión y Redes Neuronales usados para realizar predicciones [4]; Modelos Bayesianos, Machine Learning, Deep Learning y técnicas como Random Forests y Support Vector Machines utilizados para el aprendizaje autónomo [5], [6], todo este conjunto de modelos, técnicas y herramientas engloban la Analítica de Datos.

En conclusión, la Analítica de Datos se ha convertido en una herramienta muy poderosa para la sociedad, especialmente para las empresas y el gobierno, ya que les permite recolectar y almacenar grandes cantidades de datos, realizar procesos de limpieza y exploración inicial de estos, generar modelos estadísticos predictivos y de aprendizaje autónomo, y finalmente poder realizar implantación y despliegue de estos modelos para la toma de decisiones.

REFERENCIAS

- [1] Stigler, S.M. () The History of Statistics: The Measurement of Uncertainty Before 1900. Harvard University Press, Cambridge, 1986.
- [2] Hacking, I. The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference, 2nd ed. Cambridge: Cambridge University Press, 2006.
- [3] Wickham, H. and Grommum, G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. "O'Reilly Media, Inc.", 2016.
- [4] Provost, F. and Fawcett T. Data Science for Business. O'Reilly Media, Inc. 2013.
- [5] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). Bayesian Data Analysis (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>.
- [6] Goodfellow, I., Bengio, Y. and Courville A. Deep Learning. MIT Press. 2016.



Juan Luis Arias Vargas: Es Ingeniero Industrial de la Universidad Tecnológica de Pereira, Especialista en la Administración de la Informática Educativa de la Universidad de Santander, Magister en Enseñanza de la Matemática (área de Estadística) de la Universidad Tecnológica de Pereira. Es profesor asociado de la Facultad de Ciencias Básicas e Ingeniería de la Universidad Católica de Pereira, miembro del grupo de investigación Entre Ciencia e Ingeniería clasificado en A por el Sistema ScienTI Colombia. También es Investigador Asociado reconocido por el mismo Sistema. Sus áreas de interés son la ingeniería industrial y la estadística. <https://orcid.org/0000-0002-7997-1891>.